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

Paris, France, July 8-13, 2006

Sponsored by:

Edwin T. Jaynes International Center for Bayesian Methods and Maximum Entropy
Centre national de la recherche scientique, France
Université de Paris-sud, Orsay, France
Laboratoire des signaux et systèmes, France
Ministère de la défense, Direction de la recherche et développement, France
International Society for Bayesian Analysis

Sunday 09/07 (tutorials)

9:00	Registration & Coffee
10:00	A. M.-DJAFARI
11:00	B. LECOUTRE
12:00	lunch time
14:00	A PHORORILLE
15:00	C. CAVES
16:00	break time
16:30	A. CATICHA
17:15	M. DATCU

Monday 10/07

8:30			OPENING & REGISTRATION
9:00	Quantum Mechanics	C. M. CAVES	A. POHORILLE invited talk (32)
9:45	Quantum Mechanics	C. M. CAVES	A. CATICHA (48/49)
10:15	Quantum Mechanics	C. M. CAVES	L. F. Lemmens (57)
10:45	Quantum Mechanics		break time
11:15	Quantum Mechanics	A. CATICHA	C. M. CAVES invited talk (93)
12:00	Quantum Mechanics	A. CATICHA	A. VOUDRAS (35)
12:30	Quantum Mechanics	A. CATICHA	Poster Presentations
13:00			lunch time
14:00	Poster Session		Quantom Mech., Inverse Pblms. & Source Sep. (U. v. TOUSSAINT)
15:00	Source Separation	K. KNUTH	C. F. CAIAFA (10)
15:30	Source Separation	K. KNUTH	Y. NISHIMORI (87)
16:00	Source Separation	K. KNUTH	F. VRINS (81)
16:30	Source Separation		break time
17:00	Source Separation	C. JUTTEN	S. HOSSEINI (58)
17:30	Source Separation	C. JUTTEN	S. MOUSSAOUI (83)
18:00	Source Separation	C. JUTTEN	M. BABIE-ZADEH (82)

Tuesday 11/07

8:45	Information Geometry & Bayesian Nets	H. SNOUSSI	S. FIORI invited talk (92)
9:30	Information Geometry & Bayesian Nets	H. SNOUSSI	C. C. RODRIGUEZ (37)
10:00	Information Geometry & Bayesian Nets	H. SNOUSSI	A. RAMER (100)
10:30	Information Geometry & Bayesian Nets		break time
11:00	Information Geometry & Bayesian Nets	C. RODRIGUEZ	N. CATICHA (15)
11:30	Information Geometry & Bayesian Nets	C. RODRIGUEZ	D. E. HOLMES (6)
12:00	Information Geometry & Bayesian Nets	C. RODRIGUEZ	E. BJÖRNEMO (63)
12:30	Information Geometry & Bayesian Nets	C. RODRIGUEZ	Poster Presentations
13:00			lunch time
14:00	Poster Session		Bayesian Networks & Inverse Problems (R. FISCHER)
15:00	Inverse Problems	R. PREUSS	K. M. HANSON (72)
15:30	Inverse Problems	R. PREUSS	R. FISCHER (66)
16:00	Inverse Problems	R. PREUSS	U. v. TOUSSAINT (52)
16:30	Inverse Problems		break time
17:00	Inverse Problems	K. M. HANSON	R. PREUSS (76)
17:30	Inverse Problems	K. M. HANSON	V. MAZET (17)
			The "Diner Banket" (19h30 - 23h30)

Wednesday 12/07

8:45	Bayesian Inference & Image Processing	J. CENTER	Z. GHAHRAMANI invited talk (74)
9:30	Bayesian Inference & Image Processing	J. CENTER	K. KNUTH (55)
10:00	Bayesian Inference & Image Processing	J. CENTER	J. SKILLING (67)
10:30	Bayesian Inference & Image Processing		break time
11:00	Bayesian Inference & Image Processing	J. SKILLING	W. PIECZYNSKI (21)
11:30	Bayesian Inference & Image Processing	J. SKILLING	G. DELYON (18)
12:00	Bayesian Inference & Image Processing	J. SKILLING	M. SOCCORSI (12)
12:30	Bayesian Inference & Image Processing	J. SKILLING	Poster Presentations
13:00			lunch time
14:00	Poster Session		Applications & Image Processing (Z. GHAHRAMANI)
15:00			Group Photo
16:00			Committee Meeting

Thursday 13/07

8:45	Entropy & Data Processing	M. GRENDAR	V. GIRARDIN (38)
9:15	Entropy & Data Processing	M. GRENDAR	J. F. BERCHER (23)
9:45	Entropy & Data Processing	M. GRENDAR	E. V. VAKARIN (64)
10:15	Entropy & Data Processing		break time
10:45	Entropy & Data Processing	V. GIRARDIN	A. ZARZO (75)
11:15	Entropy & Data Processing	V. GIRARDIN	A. SOLANA-ORTEGA (89)
11:45	Entropy & Data Processing	V. GIRARDIN	P. L. N. INVERARDI (9)
12:15	Entropy & Data Processing	V. GIRARDIN	M. GRENDAR (85)
12:45			lunch time
14:00	Applications	E. BARAT	J. M. STERN (68)
14:30	Applications	E. BARAT	J. WELCH (71)
15:00	Applications	E. BARAT	M. JOHANSSON (51)
15:30	Applications		break time
16:00	Applications	J. M. STERN	E. BARAT (24)
16:30	Applications	J. M. STERN	F. DESBOUVRIES (39)
17:00	Applications	J. M. STERN	Tj. R. BONTEKOE (29)

Allocated times (min)	
invited talk	0:45
talk	0:30
break	0:30
lunch	1:00
poster session	1:00

MaxEnt 2006 Program Information

Saturday July 8, 2006: Welcome reception (Cité Internationale de Paris)

14:00	16:00	Registration
16:00	18h00	Welcome reception

Sunday July 9, 2006, Tutorial day (Cité Internationale de Paris)

09h00	10h00	Registration and Coffee
10h00	11h00	Tutorial 1: Ali Mohammad-Djafari: Maximum Entropy and Bayesian inference: Where do we stand and where do we go ?
11h00	12h00	Tutorial 2: Bruno Lecoutre: And if you were a Bayesian without knowing it ?
12h00	14h00	Lunch
14h00	15h00	Tutorial 3: Andrew Pohorille: Exploring the connection between sampling problems in Bayesian inference and statistical mechanics
15h00	16h00	Tutorial 4: Carlton Caves: Introduction to quantum computation
16h00	16h15	Break
16h15	17h00	Tutorial 5: A. Caticha: Updating probabilities
17h00	17h45	Tutorial 6: Mihai Datcu: Information theory based inference in the Bayesian context: applications for semantic image coding
18h00	22h00	Footbal World cup: Walk around Paris coffees and share the footbal excitings.

Monday July 10, 2006, (CNRS, Paris)

8h00	8h30	Registration
8h30	9h00	Opening and official talks

Oral session 1: Information, Probability, Quantum systems (Chair: C. Caves)

09h00	09h45	Invited talk 1: A. Pohorille: A Bayesian approach to calculating free energies of chemical and biological systems
09h45	10h15	A. Caticha From objective amplitudes to Bayesian probabilities
10h15	10h45	L.F. Lemmens Probability assignment in a quantum statistical model
10h45	11h15	Break

Oral session 2: Bayesian Probability, Quantum systems (Chair: A. Caticha)

11h15	12h00	Invited Talk 2: C. Caves Why We Should Think of Quantum Probabilities as Bayesian Probabilities
12h00	12h30	A. Vourdas : Phase space methods in continuous tensor products of Hilbert spaces
12h30	13h00	One minute Poster Session 1 presentation (Chair: U.V. Toussaint)
13h00	14h00	Lunch

14h00	15h00	**Poster session 1: (Chair: U.V. Toussaint)** 003_Bityukov, 004_Barkova, 042_Lin, 046_Gadjiev, 050_Bastami, 065_Yanez, 079_Amiri-Sazi, 095_Olmos, 096_Lopez-Rosa, 019_Mohtashami, 041_Finn, 053_Yari, 090_Pappalardo.

Oral session 3: Source Separation (Chair: Ch. Jutten)

15h00	15h30	C.F. Caiafa : A minimax entropy method for blind separation of dependent components in astrophysical images
15h30	16h00	Y. Nishimori : Riemannian optimization method on the generalized flag manifold for complex and subspace ICA
16h00	16h30	F. Vrins : Electrode selection for non-invasive Fetal Electrocardiogram Extraction using Mutual Information Criteria
16h30	17h00	Break

Oral session 4: Source Separation (Chair: K. Knuth)

17h00	17h30	Sh. Hosseini: Maximum likelihood separation of spatially auto-correlated images using a Markov model
17h30	18h00	S. Moussaoui Mars Hyperspectral Data Processing using ICA and Bayesian Positive Source Separation
18h00	18h30	M. Babaie-zadeh A fast method for sparse component analysis based on iterative detection-projection

Tuesday July 11, 2006: (CNRS, Paris)

Oral session 5: Information Geometry and Bayesian nets (Chair: H. Snoussi)

8h30	9h15	Invited talk 3: S. Fiori: Extrinsic geometrical methods for neural blind deconvolution
9h15	9h45	C. Rodriguez : Data, Virtual Data, and Anti-Data
9h45	10h15	A. Ramer : GraphMaxEnt
10h15	10h45	Break

Oral session 6: Information Geometry - Bayesian nets (Chair: C. Rodriguez)

10h45	11h15	N. Caticha : The evolution of learning systems: to Bayes or not to be
11h15	11h45	D.E. Holmes : Determining Missing Constraint Values in Bayesian Networks with Maximum Entropy: A First Step Towards a Generalized Bayesian Network
11h45	12h15	E. Bjornemo : Sensor network node scheduling for estimation of a continuous field
12h30	13h00	One minute Poster Session 2 presentation (Chair: R. Fischer)
13h00	14h00	Lunch
14h00	**15h00**	**Poster session 2: (Chair: R. Fischer)** 002_Borges, 007_Dobrzynski, 011_Barbaresco, 016_Center, 022_Bercher, 025_Dautremer, 026_Niven, 031_Alamino, 034_Neisy, 043_Cafaro, 044_Dodt, 045_Dreier, 047_Krajsek, 061_Mohammad-Djafari, 063_Bjornemo, 073_Sahmoodi, 077_Costache, 078_Esmer, 080_Kiss, 084_Roy, 084_Karimi, 091_Snoussi.

Oral session 7: Inverse problems (Chair: R. Preuss)

15h00	15h30	K.M. Hanson : Probing the covariance matrix
15h30	16h00	R. Fischer : Integrated data analysis: non-parametric profile gradient estimation
16h00	16h30	U.V. Toussaint : Parameter estimation of ellipsometry measurements
16h30	17h00	Break

Oral session 8: Inverse problems (Chair: K. Hanson)

17h00	17h30	R. Preuss Bayesian analysis on plasma confinement data bases
17h30	18h00	V. Mazet, D. Brie and J. Idier : Decomposition of a chemical spectrum using a marked point process and a constant dimension model
19h30	23h30	**Conference Dinner**

Wednesday July 12, 2006, (CNRS, Paris)

Oral session 9: Bayesian inference and Image processing (Chair: J. Center)

8h30	9h15	Invited talk 4: Z. Ghahramani A Bayesian approach to information retrieval from sets of items
9h15	9h45	K. Knuth Clearing up the mysteries: computing on hypothesis spaces
9h45	10h15	J. Skilling: Calibration and interpolation
10h15	10h45	Break

Oral session 10: Bayesian inference - Image processing (Chair: J. Skilling)

10h45	11h15	W. Pieczynski : Unsupervised segmentation of hidden semi-Markov non stationary chains
11h15	11h45	G. Deylon : Minimal stochastic complexity image partionning with non parametric statistical model
11h45	12h15	M. Soccorsi : Space-Variant Model Fitting and Selection for Image Denoising and Information Extraction
12h15	13h00	One minute Poster Session 3 presentation (Chair: Z. Gharamani)
13h00	14h00	Lunch
14h00	15h00	**Poster session 3: (Chair: Z. Gharamani)** 001_Kyo, 005_Khireddine, 008_Zarpak, 013_Gueguen, 014_Chaabouni, 020_Jalobeanu, 028_Roemer, 030_Goggens, 040_Desbouvries, 054_Amintoosi, 062_Aronsson, 069_Aminghafari, 070_Mehmood, 071_Verdoolaege, 086_Abrishami, 097_Mohammadpour.
15h00	15h30	**Group Photo (Chair: R. Bontekoe)**
15h30	16h30	**Committee meeting (Chair: K. Knuth)**

Free time

Thursday July 13, 2006, (CNRS, Paris)

Oral session 11: Entropy and Data Processing (Chair: M. Grendar)

8h45	9h15	V. Girardin : Entropy and semi-Markov processes
9h15	9h45	J.F. Bercher : An amended MaxEnt formulation for deriving Tsallis factors, and associated issues
9h45	10h15	E.V. Vakarin : Maximum entropy approach to characterization of random media
10h15	10h45	Break

Oral session 12: Entropy and Data Processing (Chair: V. Girardin)

10h45	11h15	A. Zarzo : The minimum cross-entropy method: a general algorithm for one-dimensional problems
11h15	11h45	A. Solana-Ortega : Entropic inference for assigning probabilities: some difficulties in axiomatics and applications
11h45	12h15	M. Grendar : Empirical maximum entropy methods
12h15	12h45	P.L.N Inverardi : A New Bound for Discrete Distributions based on Maximum Entropy
13h00	14h00	**Lunch**

Oral session 13: Entropy, Bayes and Applications (Chair: E. Barrat)

14h00	14h30	J.M. Stern : The Full Bayesian Significance Test for Separate Hypotheses
14h30	15h00	J. Welch : Comparing Class Scores in GCSE Modular Science
15h00	15h30	M. Johansson : Competitive bidding in a certain class of auctions
15h30	16h00	Break

Oral session 14: Entropy, Bayes and Applications (Chair: J.M. Stern)

16h00	16h30	E. Barat : Nonparametric Bayesian estimation of x/gamma-ray spectra using a hierarchical Polya tree -- Dirichlet mixture model
16h30	17h00	F. Desbouvries : Entropy computation in partially observed Markov chains
17h00	17h30	Tj. R. Bontekoe : Scheduling of schools
17h30	18h30	**Ending session (Chair: Local organizers)**

List of authors

B. Abbasi Bastami: p. 106
H. Abrishami: p. 174
B. Ait-el-Fquih: p. 86
S. Akaho: p. 176
R.C. Alamino: p. 70
S.A. Ali: p. 92
A.A. Amini: p. 166
M.G. Amin: p. 148
H. Amindavar: p. 106
M. Aminghafari: p. 140
M. Amiri-Sazi: p. 160
J.C. Angulo: p. 192
D. Aronsson: p. 128, 126

M. Babaizadeh: p. 166
B. Bakar: p. 118
E. Barat: p. 60, 58
F. Barbaresco: p. 32
E.A. Barkova: p. 18
K. Benmahammed: p. 20
J.-F. Bercher: p. 56, 54
W. Biel: p. 144
S.I. Bityukov: p. 18, 16
E. Björnemo: p. 126, 128
T. R. Bontekoe: p. 66
W. Borges: p. 14
D. Brie: p. 44

C. Cafaro: p. 92
C. Caiafa: p. 30
A. Caticha: p. 104, 102
N. Caticha: p. 70, 40
C.M. Caves: p. 188, 186
J. Center: p. 42
H. Chaabouni: p. 38
Y. Chi: p. 68
M. Costache: p. 156
D. Cowling: p. 64
J.C. Cuchi: p. 152

M. Datcu: p. 156, 38, 36, 34, 198
T. Dautremer: p. 60, 58
J.S. Dehesa: p. 132, 190, 192
V. Dehghan: p. 174
G. Delyon: p. 46
F. Desbouvries: p. 84, 86
Y. Deville: p. 120
A. Dinklage: p. 134, 154, 96, 94
L. Dobrzynski: p. 24
D. Dodt: p. 94
V. Dose: p. 134, 154
H. Dreier: p. 96
B. Duriez: p. 54

A.E. Ekimov: p. 142
O. Esmer: p. 158

S.R. de Faria Jr.: p. 138
R. Farnoosh: p. 26
L.S. Finn: p. 88
S. Fiori: p. 184
R. Fischer: p. 96, 134, 94

B.R. Gadgjiev: p. 98
F. Galland: p. 46
Z. Ghahramani: p. 150
A. Giffin: p. 102, 92
V. Girardin: p. 82
P.M. Goggans: p. 142, 68
S. Goodarzi: p. 48
M. Grendar: p. 172
L. Gueguen: p. 36
R. Guidara: p. 120
H. Gzyl: p. 28

K.M. Hanson: p. 146
H. Hauksdottir: p. 168
K.A. Heller: p. 150
C. Hérail: p. 164
M. Hirsch: p. 96

D.E. Holmes: p. 22
S. Hosseini: p. 120

J. Idier: p. 44
P.L.N. Inverardi: p. 28

A. Jalobeanu: p. 50
P. Jardin: p. 54
R. Jaspers: p. 144
M. Johansson: p. 128, 126, 108
G. Judge: p. 172
C. Jutten: p. 120, 166, 168, 164

A. Khireddine: p. 20
E. Kiss: p. 162
K. Knuth: p. 114
K. Krajsek: p. 100
N.V. Krasnikov: p. 16
E. Kuruoglu: p. 30
K. Kyo: p. 12

J. Lapuyade-Lahorgue: p. 52
M.S. Lauretto: p. 138
B. Lecoutre: p. 116
C. Lemen: p. 36
L. F. Lemmens: p. 118
S.-K. Lin: p. 90
S. Lopez-Rosa: p. 192

H. Maitre: p. 156
V. Mazet: p. 44
A. Mehmood: p. 142
R. Mester: p. 100
A. Mohammad-Djafari: p. 182, 194, 196, 124
A. Mohammadpour: p. 194
G.R. Mohtashami Borzadaran: p. 48
G. Mohtashami: p. 112
S. Moussaoui: p. 168

A. Neisy: p. 76
Y. Nishimori: p. 176
R.K. Niven: p. 62
H. Noda: p. 12

B. Olmos: p. 190

M. Pappalardo: p. 180
F. Parmentier: p. 164
B.B. Pereira: p. 138
W. Pieczynski: p. 52
M.D. Plumbley: p. 176
A. Pohorille: p. 72, 74
R. Preuss: p. 154
A. Proto: p. 30

M. Quartulli: p. 34

A. Ramer: p. 200
P. Réfrégié: p. 46
C. Rodriguez: p. 80
L. Roemer: p. 64
D. Roy: p. 170

J.M. Sabatier: p. 142
M. Sahmoodi: p. 148
R. Sameni: p. 164
D. Satua: p. 24
F. Schmidt: p. 168
J. Skilling: p. 136
V.V. Smirnova: p. 16
H. Snoussi: p. 182
M. Soccorsi: p. 34
A. Solana-Ortega: p. 178
V. Solana: p. 178
J.M. Stern: p. 138
J.M. Stern: p. 14
K. Szymaski: p. 24

A. Tagliani: p. 28
V.A. Taperechkina: p. 18
C. Tison: p. 38
U.v. Toussaint: p. 110
F. Tupin: p. 38
T. Trigano: p. 60

E.V. Vakarin: p. 130
G. Verdoolaege: p. 144
M. Verleysen: p. 164
V. Vigneron: p. 164
F. Vrins: p. 164
M.G. Von Hellermann: p. 144
A. Vourdas: p. 78

J. Welch: p. 122

R.J. Yanez: p. 132, 190
G. Yari: p. 112, 26

B. Zarpak: p. 26
A. Zarzo: p. 152

BAYESIAN ANALYSIS OF CROSS-PREFECTURE PRODUCTION FUNCTION WITH TIME VARYING STRUCTURE IN JAPAN

Koki Kyo* and Hideo Noda
Asahikawa University
3-23 Nagayama, Asahikawa, Hokkaido 079-8501, Japan

Abstract
The objective of this paper is to examine the performance of post-war Japanese economy using a production function of economic growth model. The basic framework is a variation of aggregate production function used by Solow (1956), Mankiw, Romer, and Weil (1992), etc. We consider the Cobb=Douglas production function with private capital, public capital, human capital and labour as inputs, so production for prefecture i at time t is represented by

$$Q_i(t) = K_i(t)^{\alpha_i} G_i(t)^{\beta_i} H_i(t)^{\gamma_i} [A_i(t) L_i(t)]^{1-\alpha_i-\beta_i-\gamma_i} \quad (i = 1, 2, \ldots, m),$$

where $Q_i(t)$ is output, $K_i(t)$ is the stock of private capital, $G_i(t)$ is the stock of public capital, $H_i(t)$ is the stock of human capital, $L_i(t)$ is the size of the labour force and $A_i(t)$ is a productivity index which summarizes the level of technology. The above model can be expressed in a form of linear model under the logarithmic tranformation. A set of Bayesian models is constructed by using smoothness priors for values related to $A_i(t)$ and non-informative priors for the parameters α_i, β_i and γ_i. Furthermore, Monte Carlo filter and smoother approach is applied to estimate the parameters. We show the effects of the private capital, the public capital and the human capital on output by analyzing the values of these parameters. The related result was firstly reported by Kyo and Noda (2005).

References:

[1] K. Kyo, and H. Noda (2005), Statistical analysis of cross-prefecture production function with dynamic structure in Japan, *Paper Presented at International Symposium: Intersection, Fusion and Development of Multi-Fields*, Chinese Academy of Science and Engineering in Japan.

[2] Mankiw, N. G., D. Romer, and D. Weil (1992), A Contribution to the Empirics of Economic Growth, *Quarterly Journal of Economics,* Vol.107, pp.407-437.

[3] Solow, R. M. (1956), A Contribution to the Theory of Economic Growth, *Quarterly Journal of Economics,* Vol.70, pp.65-94.

Key Words: PRODUCTION FUNCTION, SMOOTHNESS PRIORS, MONTE CARLO FILTER AND SMOOTHER

*E-mail: x.q.jiang@m2.dion.ne.jp FAX: +81-166-488718

FBST: Compositionality

Wagner Borges* and Julio M. Stern†

*Mackenzie Presbiterian University, wborges@mackenzie.com.br
†University of São Paulo, jstern@ime.usp.br

Abstract. In this paper, the relationship between the credibility of a complex hypothesis, H, and those of its constituent elementary hypotheses, H^j, $j = 1 \ldots k$, is analyzed, in the independent setup, under the Full Bayesian Significance Testing (FBST) mathematical apparatus.
Key words: Bayesian models; Complex hypotheses; Compositionality; Mellin convolution; Possibilistic and probabilistic reasoning; Significance tests; Truth values, functions and operations.

INTRODUCTION

The Full Bayesian Significance Test (FBST) has been introduced by Pereira and Stern (1999), as a coherent Bayesian significance test for sharp hypotheses. For detailed definitions, interpretations, implementation and applications, see the authors' previous articles, including two papers in this conference series, [9], [17].

In this paper we analyze the relationship between the credibility, or truth value, of a complex hypothesis, H, and those of its elementary constituents, H^j, $j = 1 \ldots k$. This problem is known as the question of *Compositionality*, which plays a central role in analytical philosophy, see [3].

According to Wittgenstein [22], (2.0201, 5.0, 5.32):

- Every complex statement can be analyzed from its elementary constituents.

- Truth values of elementary statement are the results of those statements' truth-functions (Wahrheitsfunktionen).

- All truth-function are results of successive applications to elementary constituents of a finite number of truth-operations (Wahrheitsoperationen).

The compositionality question also plays a central role in far more concrete contexts, like that of reliability engineering, see [1] and [2], (1.4):

"One of the main purposes of a mathematical theory of reliability is to develop means by which one can evaluate the reliability of a structure when the reliability of its components are known. The present study will be concerned with this kind of mathematical development. It will be necessary for this purpose to rephrase our intuitive concepts of structure, component, reliability, etc. in more formal language, to restate carefully our assumptions, and to introduce an appropriate mathematical apparatus."

When brought into a parametric statistical hypothesis testing context, a complex hypothetical scenario or complex hypothesis is a statement, H, concerning $\theta = (\theta^1, \ldots, \theta^k) \in \Theta = (\Theta^1 \times \ldots \times \Theta^k)$ which is equivalent to a logical composition of statements, $H^1, \ldots, H^k$, concerning the elementary components, $\theta^1 \in \Theta^1, \ldots, \theta^k \in \Theta^k$, respectively. Within this setting, means to evaluate the credibility of H, as well as that of

THE TRANSFORM BETWEEN THE SPACE OF OBSERVED VALUES AND THE SPACE OF POSSIBLE VALUES OF THE PARAMETER

Sergey I. Bityukov[1], Nikolai V. Krasnikov[2], Vera V. Smirnova[1]
(1) Institute for high energy physics, 142281 Protvino, Russia
(2) Institute for nuclear research RAS, Moscow, Russia
(e-mail: Serguei.Bitioukov@cern.ch, http://home.cern.ch/bityukov)

Abstract

In ref [1] the notion of statistically dual distributions is introduced. The reconstruction of confidence density [2] for the location parameter for several pairs of statistically dual distributions (Poisson and Gamma, normal and normal, Cauchy and Cauchy, Laplace and Laplace) in the case of single observation of the random variable is an unique. It allows to introduce the Transform between the space of observed values and the space of possible values of the parameter.

References:

[1] S.I. Bityukov et al. e-Print: math.ST/0411462; S.I. Bityukov et al. Proc. of Conf. PhyStat'05, http://www.physics.ox.ac.uk/phystat05/proceedings/

[2] S.I. Bityukov et al., AIP Conf.Proc. 803 (2005) 398-402

Key Words: Uncertainty, Measurement, Estimation

LAPLACE DISTRIBUTIONS AS CONJUGATE FAMILIES

E.A. Barkova[1], S.I. Bityukov[2], V.A. Taperechkina[1]
(1) Moscow State Academy of Instrument Engineering and Computer Science, Moscow, Russia
(2) Institute for high energy physics, 142281 Protvino, Russia
(e-mail: Serguei.Bitioukov@cern.ch, http://home.cern.ch/bityukov)

Abstract

In ref. [1] is introduced the notion statistically dual distributions and is shown that several pairs of distributions (Poisson and Gamma, normal and normal, Cauchy and Cauchy) are statistically dual distributions. These distributions allow to exchange the parameter and the random variable, conserving the same formula for the distribution of probabilities. The interrelation between the statistically dual distributions and conjugate families is considered in ref. [2]. It allows to use the statistical duality for estimation of the distribution parameter. In the report we show that Laplace distributions are statistically dual distributions and, correspondingly, can belong to conjugate families. The Monte Carlo experiment confirms this supposition.

References:

[1] S.I. Bityukov et al. e-Print: math.ST/0411462; S.I. Bityukov et al. Proc. of Conf. PhyStat'05, http://www.physics.ox.ac.uk/phystat05/proceedings/

[2] S.I. Bityukov et al., AIP Conf.Proc. 803 (2005) 398-402

Key Words: Uncertainty, Measurement, Estimation

WIENER FILTER IN TWO DIMENSIONNAL CASE APPLIED TO RESTORDED IMAGES

A/khireddine[1], K/Benmahammed[2]
[1] Faculty of sciences, University of Bejaia(06000), Algeria
[2] Faculty of sciences, University of Setif(19000), Algeria

ABSTRACT :

In high frequency propagation, the design of a two-dimensional Wiener filter is based on the principle of orthogonality, while being based on three following assumptions:
1. the filter used is linear and invariant.
2. the desired exit and input signal X(m,n) Z(m,n) are jointly stationary.
3. the criterion of minimization used is that of the minimal average quadratic error between the desired exit Z(m,n) and the current exit Y(m,n).

The filter of two-dimensional Wiener is a generalization of the filter of unidimensional Wiener.
Modern digital technology has made it possible to manipulate multi-dimensional signals with systems that range from simple digital circuits to advanced parallel computers[1,2]. The theory of Wiener gives the filter which minimizes the residual error (difference between the real exit and the desired exit), thus, the filter of Wiener 2D gives a solution to many problems of two-dimensional signal processing such as the restoration of degraded images. However, since the determination of this filter implies the solution of a linear equations system with great dimension, fast algorithms are necessary. The effort of calculation for the determination of the coefficients of this filter depends primarily on the statistical nature of the input signal.
Further, we will restrict ourselves to two-dimensional (2D) image processing although most of the concepts and techniques that are to be described can be extended easily to three or more dimensions. The Wiener filter is a solution to the restoration problem based upon the hypothesized use of a linear filter and the minimum mean-square (or mms) error criterion. In the example given below the image a[m,n] was distorted by a bandpass filter and then white noise was added to achieve an Signal/noise ratio equal to 30 dB[3,4].

KEY WORDS :
Digital images, Fourier transform, sampling, Wiener filter, noise.

Determining Missing Constraint Values in Bayesian Networks with Maximum Entropy: A First Step Towards a Generalized Bayesian Network

Dawn E. Holmes

Department of Statistics and Applied Probability, South Hall,
University of California, Santa Barbara,
CA 93106, USA.

Abstract. The author's past work in this area has shown that the probability of a state of a Bayesian network, found using the standard Bayesian techniques, could be equated to the Maximum Entropy solution and that this result enabled us to find minimally prejudiced estimates of missing information in Bayesian networks. In this paper we show that in the class of Bayesian networks known as Bayesian trees, we are able to determine missing constraint values optimally without the use of Bayesian techniques, using only the Maximum Entropy Formalism. We also show that it is possible to produce a generalized Bayesian network, which is specified entirely within the Maximum Entropy formalism.

Keywords: Bayesian networks, Maximum entropy, *d*-separation.
PACS: 02.50.Cw, 89.70.+c , 05.70.–a, 65.40.Gr

RECONSTRUCTION OF HYPERFINE FIELDS DISTRIBUTIONS BY THE MAXIMUM ENTROPY METHOD

Ludwik Dobrzynski[1,2], Dariusz Satuna[1], Krzysztof Szymanski[1]
(1) Institute of Experimental Physics,
University of Bialystok, Lipowa 41, 15-424 Bialystok, Poland
(2) The Soltan Institute for Nuclear Studies,
05-400 Otwock-Swierk, Poland

Abstract

The reconstruction of the hyperfine field distributions from the Mossbauer spectra is a difficult task because one tries to retrieve information about a 2- or 3-dimensional object from a single one-dimensional spectrum. In conventionally accepted procedures one has to make many simplifying assumptions in order to get this kind of information. In particular one has to assume the existence of certain correlations between the parameters: the intensity of the hyperfine field, B, quadrupole splitting, QS, and isomer shift, IS. In our paper [1] it was shown that one can successfully obtain the (B,IS) distributions even when a uniform prior is used. However, as demonstrated hereafter, this task turns out to be more difficult in the case of paramagnets, for which QS and IS distributions only can be considered. In this case one deals with a great many possible solutions and MaxEnt algorithm is not selecting the intuitively expected one. This so-called ambiguity problem can be solved only when a non-uniform prior is used. This same necessity of a using non-uniform prior exists when retrieving 3-dimensional, i.e. (B,QS,IS) distributions, but we show in the present work that in both considered cases one can devise an efficient strategy and achieve physically valuable results.

1.L.Dobrzy?ski, K.Szyma?ski, D.Satu?a, "Maximum Entropy Method in Mssbauer Spectroscopy", Nukleonika 49, Suppl. 3 (2004) S89

Key Words: Entropy measures, Blind Source Separation (BSS), astrophysical images.

IMAGE SEGMENTATION USING GAUSSIAN MIXTURE MODELS

Rahman.Farnoosh, Gholamhossein.Yari, Behnam.Zarpak
Department of Applied Mathematics
University of Science and Technology
16844, Narmak, Tehran, Iran
$\{rfarnoosh, yari, zarpak\}@iust.ac.ir$

Abstract

Recently stochastic models such as mixture models, graphical models, Markov random fields and hidden Markov models are key role in probabilistic data analysis. Gaussian mixture model as generalized normal distribution is also an important tool in image data analysis. Also image segmentation means to divide one picture into different types of classes or regions, for example a picture of nature has some classes like 'sky','mountains', 'trees','river' and so on. Therefore we can suppose that each class has normal distribution with specify mean,variance and generally a picture has Gaussian mixture model. This is independent identically distribution (i.i.d) case which is well known. In this paper we have learned Gaussian mixture model to the pixels of an image as training data and the parameters of the model are learned by EM-algorithm. Meanwhile pixel labeling corresponded to each pixel of true image is done by Bayes rule. This hidden Markov image is the form of Potts Markov random field, So we can automatically classify an image with this hidden or labeled image. In addition it is natural to assume that Markov property of upper orders holds in each regions. It means that data in each class are close together with uniform means, low variations and high correlations. Thus we can do Gaussian mixture models in spatial domain that is if each class has Gaussian Markov random field so the general image is a Gaussian mixture Markov random field. We then show some experiments.

Key Words: Gaussian Mixture Model (GMM), Image Segmentation, Bayes Rule, Expectation-Maximization (EM)Algorithm, Potts Markov Random Field (PMRF), Gaussian Markov Random Field (GMRF), Gaussian Mixture Markov Random Field (GMMRF).

A NEW BOUND FOR DISCRETE DISTRIBUTIONS BASED ON MAXIMUM ENTROPY

Henryk Gzyl[1], Pier Luigi Novi Inverardi[2], Aldo Tagliani[2]
(1) Dept. of Statistics - USB
AP 890000 Caracas 1080-A (Venezuela)
(e-mail: hgzyl@reacciun.ve - Ph: +58 212 29063386)
(2) Dept. of Computer and Management Sciences - University of Trento
via Inama, 1 - 38100 Trento (Italy)
(e-mail: PierLuigi.NoviInverardi@unitn.it - Fax: +39 461 882124)
(e-mail: Aldo.Tagliani@unitn.it - Fax: +39 461 882124)

Abstract

In this paper we compare some classical and well known bounds (as Chernoff's bound or moment bounds) for nonnegative integer-valued random variables for estimating the survival probability and a new tighter bound stemming from Maximum Entropy technique constrained by fractional moments given by $\mathbb{E}(X^\alpha)$, $\alpha \in \mathbb{R}^+$.

Because the classical bounds are usually given in terms of integer moments or in terms of moment generating function, they may be able to exploit only partially the information contained in the data: for this reason these bounds are not very tight although they can be easily calculated.

We exploit a result of Lin (1992) which supports the characterization of a distribution through its fractional moments and we show (Novi Inverardi and Tagliani (2003)) that the Maximum Entropy probability mass function $P_M^{(\mathrm{fm})}$ recovered involving M fractional moments converges in entropy to the true probability mass function P. This last result means that if we are interested in approximating a discrete distribution and/or some its characteristic constants (think to expected values, tails, probabilities or other) the equivalent counterparts evaluated on $P_M^{(\mathrm{fm})}$ are as close as we like to the true values and the closeness depends on the (increasing) value of M.

But usually the available knowledge on a distribution is expressed by integer moments and not by fractional moments. This means that we need a link between moment generating function and/or integer moments and fractional moments.

Traditionally the moment generating function of a random variable X is used to generate positive integer moments of X. But it is clear that the moment generating function also contains a wealth of knowledge about arbitrary real moments and hence, on fractional moments. Taking this into account, to obtain fractional

A MINIMAX ENTROPY METHOD FOR BLIND SEPARATION OF DEPENDENT COMPONENTS IN ASTROPHYSICAL IMAGES

Cesar F. Caiafa[1], Ercan E. Kuruoğlu[2] and Araceli N. Proto[1,3]

(1) Lab. de Sistemas Complejos. Fac. de Ingeniería (UBA), Buenos Aires, Argentina

(2) Istituto di Scienza e Tecnologie dell'Informazione - CNR, Pisa, Italy

(3) Comisión de Inv. Científicas de la Prov. de Buenos Aires, Buenos Aires, Argentina

(e-mail: ccaiafa@fi.uba.ar, ercan.kuruoglu@isti.cnr.it and aproto@fi.uba.ar)

Abstract

We develop a new technique for the blind separation of potentially non independent components in astrophysical images. Given a set of linearly mixed images, corresponding to different measurement channels, we estimate the original electromagnetic radiation sources in a blind fashion. Specifically, we investigate the separation of cosmic microwave background (CMB), thermal dust and galactic synchrotron emissions without imposing any assumption on the mixing matrix. In our approach, we use the Gaussian and non-Gaussian features of astrophysical sources and we assume that CMB-dust and CMB-synchrotron are uncorrelated pairs while dust and synchrotron are correlated which is in agreement with theory. These assumptions allow us to develop an algorithm which associates the Minimum Entropy solutions with the non-Gaussian sources (thermal dust and galactic synchrotron emissions) and the Maximum Entropy solution as the only Gaussian source which is the CMB. This new method is more appropriate than ICA algorithms because independence between sources is not imposed which is a more realistic situation. We investigate several measures associated with entropy and we compare them. Finally, we present an example of separation using the Euclidean distance between the Gaussian probability density function (pdf) and a Parzen based estimation of the pdf associated with the data. For the validation of our approach we present experimental results using a database that simulates de one expected from the instruments that will operate onboard ESA's *Planck* Surveyor Satellite to measure the CMB anisotropies all over the celestial sphere.

Key Words: Entropy measures, Blind Source Separation (BSS), astrophysical images.

MaxEnt'06, Paris , July 2006

Information Intrinsic Geometric Flows : Kähler-Ricci & Calabi Flows on Siegel & Hyper-Abelian Metrics of Complex Autoregressive Models

Frédéric BARBARESCO, Thales Air Defence, 7/9 rue des Mathurins F-92223 Bagneux, France
E-mail : frederic.barbaresco@fr.thalesgroup.com , Phone : 33.(0)1.40.84.20.04

1. Preambule

Geometric Flow Theory is cross fertilized by diverse elements coming from Pure Mathematic (geometry, algebra, analyse, PDE) and Mathematical Physic (calculus of variations, General Relativity, Einstein Manifold, String Theory), but its foundation is mainly based on Riemannian Geometry, as explained by M. Berger in a recent panoramic view of this discipline [Berger], its extension to complex manifolds, the Erich Kähler's Geometry [Kähler1], vaunted for its unabated vitality by J.P. Bourguignon [Bourguignon] in [Kähler2], and Minimal Surface Theory recently synthetized by F. Hélein [Helein]. This paper would like to initiate seminal studies for applying intrinsic geometric flows in the framework of information geometry theory. More specifically, after having introduced Information metric deduced for Complex Auto-Regressive (CAR) models from Fisher Matrix (Siegel Metric and Hyper-Abelian Metric from Entropic Kähler Potential), we study asymptotic behaviour of PARCOR parameters (reflexion coefficients of CAR models) driven by intrinsic Information geometric Kähler-Ricci and Calabi flows. These Information geometric flows can be used in different contexts to define distance between CAR models interpreted as geodesics of Entropy Manifold (e.g : distance between plane curves parametrized by CAR models).

2. Siegel Metric for Complex Autoregressive Model

Chentsov has defined main axioms of Information Geometry. In this Theory, we consider families of parametric density functions $G_\Theta = \{p(./\theta):\theta\in\Theta\}$ with $\Theta = [\theta_1 \ \cdots \ \theta_n]^T$, from which we can define a Riemannian Manifold Structure by meam of Fisher Information matrix $(g_{ij}(\theta))_{ij}$:

$g_{ij}(\theta) = E_\theta\left[\dfrac{\partial \ln p(./\theta)}{\partial\theta_i}.\dfrac{\partial \ln p(./\theta)}{\partial\theta_j^*}\right]$, with the Riemannian metric : $ds^2 = \sum_{i,j=1}^{n} g_{ij}(\theta).d\theta_i.d\theta_j^*$

This metric can also be naturally introduced by a Taylor expansion of Kullback Divergence :

$$K(\theta,\tilde\theta)\big|_{\tilde\theta=\theta+d\theta} \cong K(\theta,\theta) + \left(\frac{\partial K(\theta,\tilde\theta)}{\partial\tilde\theta}\right)^+_{\tilde\theta=\theta}(\tilde\theta-\theta) + \frac{1}{2}(\tilde\theta-\theta)^+\left(\frac{\partial^2 K(\theta,\tilde\theta)}{\partial\tilde\theta\partial\tilde\theta^*}\right)(\tilde\theta-\theta) \cong \frac{1}{2}\sum_{i,j} g_{ij}(\theta)d\theta_i.d\theta_j^*$$

We demonstrate easily that this Fisher metric is equivalent to the Siegel metric, introduced by Siegel in the 60's in the framework of Symplectic Geometry. Indeed, if we consider a Complex Multivariate Gaussian Law :

$p(X/R_n,m_n) = (2\pi)^{-n}.|R_n|^{-1}.e^{-Tr[\hat R_n.R_n^{-1}]}$ with $\hat R_n = (X-m_n).(X-m_n)^+$ such that $E[\hat R_n] = R_n$

it is well-known that the Fisher Information matrix is given by : $\mathrm{g}_{\mathrm{ij}}(\theta) = -Tr[\partial_i R_n.\partial_j R_n^{-1}] + \partial_i m_n^+.R_n^{-1}.\partial_j m_n$

In the following, we will only consider random process with zero mean $m_n = E[X] = 0$, and so if we apply the following relation $R_n.R_n^{-1} = I_n \Rightarrow \partial R_n = -R_n.\partial R_n^{-1}.R_n$, the Fisher matrix is reduced to :

$g_{ij}(\theta) = Tr[(R_n.\partial_i R_n^{-1}).(R_n.\partial_j R_n^{-1})]$ with the associated Riemannian metric :

$$ds^2 = \sum_{i,j} g_{ij}(\theta).d\theta_i.d\theta_j = \mathrm{Tr}\left[\mathrm{R_n}.\left(\sum_i \partial_i R_n^{-1}.d\theta_i\right).R_n.\left(\sum_j \partial_j R_n^{-1}.d\theta_j\right)\right] \quad \text{with} \quad \mathrm{dR_n^{-1}} = \sum_k \partial_k R_n^{-1}.d\theta_k$$

We can then observe that it is completely equivalent with Siegel Metric : $\boxed{\mathrm{ds}^2 = Tr[(R_n.dR_n^{-1})^2]}$ introduced by Karl Ludwig Siegel in his book « Symplectic Geometry ». This metric is invariant under the action of the following group $(GL_n(C),.)$: $R_n \to W_n.R_n.W_n^+$, $\mathrm{W_n} \in GL_n(C)$, and geodesics are given by :

SPACE-VARIANT MODEL FITTING AND SELECTION FOR IMAGE DE-NOISING AND INFORMATION EXTRACTION

Marco Quartulli[1], Matteo Soccorsi[2], Mihai Datcu[2]
(1) Advanced Computer System, 00139 Rome, Italy
(2) German Aerospace Center, 82234 Oberpfaffenhofen, Germany

Abstract

With the growing importance of model-based signal analysis methods, the dependence of their performance on the choice of the models needs to be addressed. Bayesian theory incorporates model selection in a natural and direct way: we apply it to the space-variant choice of the best model in a given reference class in the framework of parameter estimation from noisy data. In particular, we introduce an algorithm for image information extraction and de-noising that is based on a two-level model and estimates local texture Gauss-Markov Random Field (GMRF) parameters and local GMRF model order for incomplete data. Since model selection is based on an approximate numerical computation of the evidence integral, we propose a further selection criterion based on Rate Distortion theory for a cross validation of the results. The link between Bayesian model selection and Rate Distortion is explained. Results are presented on Synthetic Aperture Radar (SAR) images.

Key Words: GMRF, Model Selection, Parameter Estimation, Rate Distortion

Analysis of Satellite Image Time Series Based on Information Bottleneck

Lionel Gueguen[1,3], Camille Lemen[1,3], Mihai Datcu[2,1]
(1) Get-Télécom Paris, TSI dept.
46 rue Barrault, 75013 Paris, France
(2) German Aerospace Center DLR, IMF
Oberpfaffenhofen, D-82234 Wessling, Germany
(3) CNES
18 avenue Edouard Belin, 31401 Toulouse, France

Abstract

Derived from Information theory, the Information Bottleneck principle[3] enables to quantify and qualify the information contained in a signal. This paper presents an algorithm based on the Information Bottleneck principle to analyze Satellite Image Time Series (SITS). The entropic method includes a parameters estimation and a model selection. This method has been applied to textural and radiometric parametric models[1, 2]. Thus, textural and radiometric information contained in SITS has been quantified. This paper presents a method to take into account the geometry information. Two approaches are presented. On one hand, each image of the SITS is segmented and the obtained regions are described by radiometric, textural and geometric models. Using the Information Bottleneck method on these models, this approach leads to a spatio-temporal characterization of the spatial regions of the SITS. On the other hand, the geometrical information is extracted first from a segmentation , then the radiometric and textural information is extracted through the Information Bottleneck method. This approach leads to a temporal characterization of the spatial regions of the SITS.

References

[1] M. Datcu, K. Seidel, and M. Walessa. Spatial Information Retrieval from Remote-Sensing Images. i. Information Theoritical Perspectives. *IEEE Transactions on Geoscience and Remote Sensing*, 36(5):1431–1445, Sept. 1998.

[2] M. Schroder, H. Rehrauer, K. Seidel, and M. Datcu. Spatial Information Retrieval from Remote-Sensing Images. ii. Gibbs-Markov Random Fileds. *IEEE Transactions on Geoscience and Remote Sensing*, 36(5):1446–1455, Sept. 1998.

RELEVANT SCATTERS CHARACTERIZATION IN SAR IMAGES

Houda Chaabouni[1], Céline Tison[2], Florence Tupin[3], Mihai Datcu[1]
(1) German Aerospace Center, Oberpfaffenhofen, D-82234 Wessling, Germany
(2) Centre National d'Etudes Spatiales, 31401 Toulouse, France
(3) Ecole Nationale Supérieure des Télécommunications, 75013 Paris, France

Abstract

Recognizing scenes in a single look meter resolution Synthetic Aperture Radar (SAR) images, requires the capability to identify relevant signal signatures in condition of variable image acquisition geometry, arbitrary objects poses and configurations. Among the methods to detect relevant scatterers in SAR images comes the internal coherence. The SAR spectrum splitted in azimuth generates a series of images which preserve high coherence only for particular object scattering. The detection of relevant scatterers can be done by correlation study or Independent Component Analysis (ICA) methods. The present article presents the state of the art for SAR internal correlation analysis and proposes further extensions using elements of inference based on information theory applied to complex valued signals. The set of azimuth looks images is analyzed using mutual information measures and an equivalent channel capacity is derived. The localization of the target requires analysis in a small image window, thus resulting in imprecise estimation of the second order statistics of the signal. For better precision, a parametric model is inferred. The method is applied to detect and characterize relevant objects in urban areas.

Key Words: SAR, Relevant Scatters, Internal Correlation, Mutual Information, Information Theory, Parametric Model

The evolution of learning systems: to Bayes or not to be

Nestor Caticha
Instituto de Física Universidade de São Paulo,
Caixa Postal 66318 CEP 05315-970 São Paulo, Brazil

Abstract

Bayesian algorithms pose a limit to the performance learning algorithms can achieve. Natural selection should guide the evolution of information processing systems towards those limits. What can we learn from this evolution and what properties do the intermediate stages have? While this question is too general to permit any answer, progress can be made by restricting the class of information processing systems under study. We present analytical and numerical results for the evolution of on-line algorithms for learning from examples for neural network classifiers, which might include or not a hidden layer. The analytical results are obtained by solving a variational problem to determine the learning algorithm that leads to maximum generalization ability. Simulations using evolutionary programming, for programs that implement learning algorithms, confirm and expand the results.

The principal result is not just that the evolution is towards a Bayesian limit and that indeed it is essentially reached. In addition we find that evolution is driven by the discovery of useful structures or combinations of variables and operators. In different runs the temporal order of the discovery of such combinations is unique. The main result is that combinations that signal the surprise brought by an example arise always before combinations that serve to gauge the performance of the learning algorithm. This latter structures can be used to implement annealing schedules. The temporal ordering can be understood analytically as well by doing the functional optimization in restricted functional spaces. We also show that there is data suggesting that the appearance of these traits also follows the same temporal ordering in biological systems.

LEARNING COMPLEX CLASSIFICATION MODELS FROM LARGE DATA SETS

Julian L. Center, Jr.
Creative Research Corporation, 385 High Plain Rd., Andover, MA 01810
(e-mail: jcenter@ieee.org)

Abstract

Taking a Bayesian approach to designing a classification algorithm, we start with a general model form with adjustable parameters and learn a posterior probability distribution for the model parameters based on a set of data samples. In many applications, classification models can become quite complex. For example, an image recognition algorithm can be based on a mixture model with many components, each with many parameters.

Unfortunately, the computations needed to determine the posterior probability distribution often become overwhelming. Since we are considering models with a large number of parameters, a large number of samples is needed to narrow the range of probable models. Because the model is complex and there are many samples, computing the likelihood of a particular model takes significant computer time. Therefore, exploring the large model-parameter space in detail becomes an intractable problem.

Of course, if the data set is large enough, the information gain will narrow the range of probable models to a very small subset of the parameter space. If we can find this subspace quickly, we can employ our computational power to adequately explore this region. However, searching for this small region can prove difficult because it is so small and because evaluating each point in the search requires evaluating the complete likelihood function.

We present a computationally feasible solution to this problem based on breaking the large data set into several smaller data subsets. We use Bayesian theory to design an algorithm for processing the subsets in stages.

For each stage, we combine a new data subset with a prior distribution that summarizes previous stages. We employ nested sampling to focus our exploration of the parameter space on high probability areas and use slice sampling to draw candidate parameter values from the prior distribution. We then use variational methods to approximate the resulting distribution on the parameter space. This approximation summarizes the results and becomes the prior distribution for the next stage.

The end result is a discrete probability distribution on model parameter space. This leads to a classification algorithm that is a "mixture of experts" combining the classification probabilities of the best models.

DECOMPOSITION OF A CHEMICAL SPECTRUM USING A MARKED POINT PROCESS AND A CONSTANT DIMENSION MODEL

Vincent Mazet[1], David Brie[1], Jérôme Idier[2]
(1) CRAN (UHP, CNRS UMR 7039), BP 239,
54506 Vandœuvre-lès-Nancy Cedex, France
(e-mail: firstname.lastname@cran.uhp-nancy.fr)
(2) IRCCyN (ÉCN, CNRS UMR 6597),
1 rue de la Noë, BP 92101, 44321 Nantes Cedex 3, France
(e-mail: jerome.idier@irccyn.ec-nantes.fr)

Abstract

We consider the problem of estimating the peak parameters in a chemical spectrum, *i.e.* their locations, amplitudes and widths. The spectrum $\mathbf{y}$ is modelled as a noisy sum of K positive Lorentzian peaks $\boldsymbol{f}$: $\mathbf{y} = \sum_{k=1}^{K} \boldsymbol{f}(\mathbf{t}_k, \mathbf{a}_k, \mathbf{s}_k) + \mathbf{e}$ where $\mathbf{t}_k$, $\mathbf{a}_k$ and $\mathbf{s}_k$ stand respectively for the location, amplitude and width of the peak k, and $\mathbf{e}$ denotes the noise and model errors. A non-supervised Bayesian approach coupled with MCMC methods is retained to solve the problem.

A marked point process provides a suitable representation for this phenomenon: it is a a finite set of objects (*i.e.* a configuration of points with some marks) lying in a bounded space, corresponding in our application to the observation space while the objects model the peaks, characterized by their locations and marks (amplitude and widths). A stochastic model for these quantities is then proposed.

But the peak number is also unknown. Numerous MCMC methods for model uncertainty have been proposed, such as the RJMCMC algorithm. Nevertheless, we propose in this paper an approach in which the dimension model is constant. Thus, the use of a Gibbs sampler is possible and natural due to the hierarchical structure of the model. The idea consists in considering an upper bound for peak number and modelling the peak occurrence by a Bernoulli distribution. However, the estimation is not straightforward because of the label switching phenomenon; we then propose a label switching method adapted to the proposed approach.

In conclusion, this approach performs better than a classical deconvolution approach where the peaks have inevitably the same width. Moreover, the input is generally modelled as a Bernoulli-Gaussian process of N points (N being the length of $\mathbf{y}$) though we consider only a signal with length K, and, obviously, $K < N$ (the peak number is less than the signal length). Therefore, there is less variables to estimate, so the estimation is better and the method faster.

MINIMAL STOCHASTIC COMPLEXITY IMAGE PARTITIONING WITH NON PARAMETRIC STATISTICAL MODEL

Guillaume Delyon, Philippe Réfrégier, Frédéric Galland
Fresnel Institute, UMR CNRS 6133, EGIM, Domaine
Universitaire de Saint-Jérôme, 13397 Marseille Cedex 20
(e-mail: {guillaume.delyon,frederic.galland,philippe.refregier}@fresnel.fr)

Abstract

Image partitioning is an important topic in computer vision and image processing and has numerous applications such as medical or military. The sensors used to obtain these images can thus be of various types and can lead to images corrupted with different noise model. It is therefore important to take into account the physical nature of the images in statistical techniques for image segmentation.

Statistical image segmentation techniques have been extensively studied since the work of Geman and Geman. In particular, Markov Random Fields (MRF) models provide an efficient regularization method but introduce parameters that cannot be easily determined automatically and which can lead to difficult optimization problems. Variational methods for image segmentation have been recently studied but lead to the same limitations as MRF approaches since parameters that cannot be easily determined automatically are present in the criterion to optimize.

The Stochastic Complexity (SC) minimization principle, introduced by Rissanen in 1978 has been early used to address the issue of order model selection. Based on information theory, this principle allows one to estimate the number of needed parameters for parametric description of observed data. This approach consists in minimizing a joint entropy which corresponds to the sum of the number of bits needed to describe the data and the number of bits needed to describe the model.

We propose a general image partitioning method based on a statistical active grid [1]. This technique allows us to estimate the probability density functions (pdf) of the grey levels of the image with step functions. It allows us to take into account the physical origin of the fluctuations of the grey levels without making hypotheses on the pdf family which describes at best the fluctuations. We demonstrate that the partitioning can be obtained by minimizing a criterion without parameter to be tuned by the user and that corresponds to a non parametric statistical approach.

References:

[1] F. Galland *et al.*, “Minimum description length synthetic aperture radar image segmentation,” *IEEE IP*, vol. 12, no. 9, pp. 995–1006, 2003.

Key Words: Image partitioning, Stochastic Complexity, Density Estimation

Aspects of Residual Information Measures for Weighted Distributions

G. R. Mohtashami Borzadaran- Samira Goodarzi
Department of Statistics
Faculty of Science
The University of Birjand
Birjand-IRAN
Email : Gmohtashami@ birjand.ac.ir
gmb1334@yahoo.com

Abstract

The concepts of weighted distributions have been introduced by Rao (1965,1985). A weighted function will be denoted by $w(x)$ and $g(x,\theta) = \frac{w(x)f(x,\theta)}{E(w(X))}$ where $E_\theta(w(X)) = \int_D w(x)dF(x)$, and $f(.,\theta)$ is the distribution of random variable X and g is the pdf of the weighted distribution..

Characterization results for the residual information measures are given here in view of the weighted distributions. We also derive relationship among residual information measures and reliability measures such as hazard rate.The residual divergence between two positive random variables are studied and finding link results relevant to information theory and reliability theory. Some examples that lead us to results related to information measures are derived for order statistics, record value, proportional hazard, proportional reversed hazard, Lorenz curve and hazard rate as special cases of weighted families.

Ebrahimi and Kirmani(1996) defined the uncertainty of residual lifetime distributions, then Asadi et. al. (2005, 2004) obtained some results related to minimum dynamic discrimination information and maximum dynamic entropy models.We obtain results concerning their relations with life distributions and information measures and give some examples for weighted families.

Some inequalities, relations and partial ordering for weighted reliability measures are also presented. A new measure of information called cumulative residual entropy in view of M Rao et. al. (2004) is defined, examples and some of its properties are obtained at the end of this paper.

References

[1] Ebrahimi, N & Kirmani S. N. U. A. (1996). A measure of discrimination between two residual life-time distributions and its applications. *Ann. Inst. Statist. Math.*, **Vol. 48, No. 2**, 257-265.

[2] Majid Asadi, Nader Ebrahimi, G. G. Hamedani and Ehsan S. Soofi (2005). Minimum dynamic discrimination information models. *J. Appl. Probab,* **42, no. 3** , 643-660

[3] Majid Asadi, Nader Ebrahimi, G. G. Hamedani and Ehsan S. Soofi (2004). Maximum dynamic entropy models. *J. Appl. Probab,* **41, no. 2** , 379-390.

[4] M. Rao, Y. Chen , Baba C. Vemuri and F. Wang (2004). Cumulative Residual Entropy: A New Measure of Information. *IEEE Transactions on Information Theory,* **50 (6),** 1220-1228.

[5] Rao, C. R. (1965). On discrete distributions arising out of methods of ascertainment, in *Classical and Contagious Discrete Distributions,* G. Patil Ed., Pergamon Press and Statistical Publishing Society, Calcutta, 320-332.

[6] Rao, C. R. (1985). Weighted distributions arising out of methods of ascertainment, in *A Celebration of Statistics,* A. C. Atkinson & S. E. Fienberg eds. Springer Verlag, New York, Chapter 24 , pp. 543-569.

MULTISOURCE DATA FUSION FOR BANDLIMITED SIGNALS: A BAYESIAN PERSPECTIVE

André Jalobeanu
LSIIT (UMR 7005 CNRS-ULP) @ ENSPS, Illkirch, France
(e-mail: jalobeanu@lsiit.u-strasbg.fr, web: lsiit-miv.u-strasbg.fr/paseo)

Abstract

We consider data fusion as the reconstruction of a single model from multiple data sources. The model is to be inferred from a number of blurred and noisy observations, possibly from different sensors under various conditions. It is all about recovering a compound object (signal+uncertainties) that best relates to the observations and contains all the useful information from the initial data set.

We wish to provide a flexible framework for bandlimited signal reconstruction from multiple data. In this paper, we focus on a general approach involving forward modeling (prior model, data acquisition) and Bayesian inference. The proposed method is valid for n-D objects (signals, images or volumes) with multidimensional spatial elements. However, for clarity reasons, both formalism and test results will be shown in 1D for single band signals. The main originality lies in seeking an object with a prescribed point spread function (psf), for which we choose a B-spline. This ensures an optimal sampling in both signal and frequency spaces, and allows for a shift invariant processing.

The model resolution, the geometric distortions, the psf and the regularity of the sampling grid can be arbitrary for each sensor. The method was designed to handle realistic Gauss+Poisson noise. Although a simple Gaussian Markov chain was used for regularization, any efficient prior model could be employed instead.

We obtained promising results in reconstructing a super-resolved signal from two blurred and noisy shifted observations. Practical applications are under development within the SpaceFusion[1] project. In astronomical imaging, we aim at a sharp, well-sampled, noise-free and possibly super-resolved image. Virtual Observatories could benefit from such a way to combine large numbers of multispectral images from various sources. In planetary imaging or remote sensing, the 3D image formation model has to be taken into account even for flat terrains. Nevertheless, this can be addressed within the same framework.

Key Words: Model-based data fusion, reconstruction, generative models, uncertainties, B-splines, super-resolution

[1]ANR grant "Projet jeunes chercheurs 2005"

Unsupervised segmentation of hidden semi-Markov non stationary chains

Wojciech Pieczynski and Jérôme Lapuyade-Lahorgue

INT/GET, Département CITI, CNRS UMR 5157
9, rue Charles Fourier, 91000 Evry, France
Wojciech.Pieczynski@int-evry.

In the classical hidden Markov chain model we have a hidden chain X, which is a Markov one, and an observed chain Y. Hidden Markov chains are widely used in numerous problems; however, in some situations they have to replaced by the more general "hidden semi-Markov chains" [1, 3], which can be seen as particular "triplet Markov chains" T=(X, U, Y), where the auxiliary chain U models the fact that X is semi-Markov [5]. Otherwise, it has been showed that a non stationary classical hidden Markov chain can also be considered as a triplet Markov stationary chain with, as a consequence, the possibility of parameters estimation [2].
The aim of this paper is to use the both properties simultaneously. We first consider a triplet Markov chain T1=(X, U1, Y), which is equivalent to a hidden semi-Markov chain. We then consider that T1 is not stationary, which is modelled by an another stationary triplet Markov chain T2=(X, U1, U2, Y) (in T2 the auxiliary chain is U=(U1, U2)). Finally, T2 is used to estimate the hidden semi-Markov non stationary chain in an unsupervised manner. "Unsupervised" means that all the model parameters are estimated from the only observed data by an original estimator, which is a new variant of the general "iterative conditional estimation" (ICE) method [5].
We present different experiments showing the interest of the new model and related processing with respect to the classical stationary hidden semi-Markov chains.

References

[1] S. Faisan, L. Thoraval, J.-P. Armspach, M.-N. Metz-Lutz, and F. Heitz, Unsupervised learning and mapping of active brain functional MRI signals based on hidden semi-Markov event sequence models, *IEEE Trans. on Medical Imaging*, Vol. 24, No. 2, pp. 263-276, 2005.
[2] P. Lanchantin and W. Pieczynski, Unsupervised non stationary image segmentation using triplet Markov chains, Advanced Concepts for Intelligent Vision Systems (ACVIS 04), Aug. 31-Sept. 3, Brussels, Belgium, 2004.
[3] M. D. Moore and M. I. Savic, Speech reconstruction using a generalized HSMM (GHSMM), *Digital Signal Processing*, Vol. 14, No. 1, pp. 37-53, 2004.
[4] W. Pieczynski, Chaînes de Markov Triplet, *Comptes Rendus de l'Académie des Sciences – Mathématique*, Série I, Vol. 335, Issue 3, pp. 275-278, 2002.
[5] W. Pieczynski and F. Desbouvries, On triplet Markov chains, International Symposium on Applied Stochastic Models and Data Analysis, (ASMDA 2005), Brest, France, May 2005.

Bayesian classification and entropy for promoter prediction in human DNA sequences

J.-F. Bercher (1), P. Jardin (1), B. Duriez (2)
(1) Équipe Signal et Information, ESIEE, France.
(2) INSERM-U654, Molecular and cellular bases of genetic diseases, Créteil, France.

February 27, 2006

There is now a large amount of genomic data available in databases for researchers. Computational methods are yet available for data retrieval and analysis, including sequences similarity searches, structural and functionnal predictions. Computationnal detection of genes has received an important interest and many accurate methods are available. However, other functionnal sites are more difficult to characterize.
In this work, we examine the potential of entropy and bayesian tools for promoter localization in human DNA sequences. Promoters are regulatory regions (at least one for each gene, located near the first exon) that governs the expression of genes, and their prediction is reputed difficult, so that this issue is still open.
To process DNA sequences it is useful to convert them using numerical representation that preserve their statistical properties. We choose the Chaos Game representation (CGR) [Jeffrey1990] of DNA sequences which has interesting properties: the source sequence can be recovered uniquely from the CGR transcription and the distance between CGR position measures similarity between corresponding sequences. This representation is applied to sequences of “words” of variable length (number of elementary bases). Typically we used words from 1 to 6 nucleotides. Using this CGR we have put in evidence the non stationarity of the genome: coding, promoter or genomic regions of DNA result in different CGR matrices. In particular we observe the fractal depletion in CG for genomic regions (that is under-representation of CG words) and CG "islands" in about 80% of promoters.
In order to analyse DNA sequences, references probabilities of the genomic, coding and promoters background are built using data from public databases. We also estimate “local” probability distribution functions, using a sliding window, and a forgetting factor.
We built a naïve bayesian classifier for promoter detection, by testing the likelihood ratio promoter/genomic or promoter/coding of the sequence at hand. Results show that performance is interesting when the window is located near the TSS – *Transcription Start Site*, and the window length is less than 200 bases. Such a classifier has already be useful for classifying species as in [Sandberg2001].
Local probabilities were used to evaluate (i) the local entropy of the sequence, (ii) the Kullback divergence to the background (with respect to the hypothesis on the nature – genomic or promoter, of the background). Again, our experiments showed that these indicators clearly reveal the core-promoter and TSS positions in many cases. However, we also noticed, as was already pointed in the litterature [Hannenhalli2001,Zhang2003], that the set of promoters can be divided in (at least) two classes, the first one (with high CG ratio) being relatively easy to predict, while the second (that may in fact be divided in more subclasses) gives more mitigated results.
An interesting point is that a promoter prediction tool can assess or infirm the bioinformatic prediction of a gene. Such examples will be presented at the conference.

References

[Jeffrey1990] H. J. Jeffrey, *Chaos Game Representation of gene structure*. Nucleic Acids Research. 18:2163-2170, 1990.
[Sandberg2001] R. Sandberg, G. Winberg, C.-I. Bränden, A. Kaske, I. Ernberg and J. Cöster, *Capturing Whole-Genome Characteristics in Short Sequences Using a Naïve Bayesian Classifier*, Genome research, Vol. 11, Issue 8, 1404-1409, August 2001.
[Zhang2003] M.Q. Zhang, *Prediction, Annotation and Analysis of Human Promoters* - CSHL Quantitative Biology Symposium 68. pp 217-225 2003.
[Hannenhalli2001] S. Hannenhalli and S. Levy, *Promoter prediction in the human genome*, Bioinformatics, vol. 17, suppl. 1, pp S90-S96, 2001.

An amended MaxEnt formulation for deriving Tsallis factors, and associated issues

J.-F. Bercher
Équipe Signal et Information, MOSIM, ESIEE, France.

February 28, 2006

The formalism of nonextensive statistical mechanics [1, 2] leads to a generalized Boltzmann factor in the form of a Tsallis distribution (or factor) that depends on an entropic index and recovers the classical Boltzmann factor as a special limit case [1]. This distribution may behave as a power law. In a wide variety of fields, experiments, numerical results and analytical derivations fairly agree with the description by a Tsallis distribution.
Tsallis' distributions (sometimes called Levy distributions) are derived by maximization of Tsallis entropy [3], under suitable constraints. However, these distributions do not coincide with those derived for the classical MaxEnt and consequently will not be justified from a probabilistic point of view, because of the uniqueness of the rate function in the large deviations theory [4, 5]. In view of the success of nonextensive statistics, there should exist a probabilistic setting that provides a justification for the maximization of Tsallis entropy. There are now several indications that results of nonextensive statistics are physically relevant for partially equilibrated or nonequilibrated systems, with a stationary state characterized by fluctuations of an intensive parameter [6, 7].

In this work I propose an amended MaxEnt formulation for systems with a displaced equilibrium, find that the relevant entropy in this setting is the Rényi entropy, interpret the mean constraints, derive the correct form of solutions, propose numerical procedures for estimating the parameters of the Tsallis factor and characterize the associated entropies.
I show that a Tsallis-like distribution can be derived as the minimizer of the Kullback-Leibler information divergence with respect to a reference distribution Q (or equivalently as the maximizer of Shannon Q-entropy), where the minimization is carried under a mean log-likelihood constraint and a (mean) observation constraint. The mean log-likelihood constraint characterizes the 'displacement' from the conventional equilibrium. This corresponds to an amended MaxEnt formulation, where one looks for an intermediate distribution between two references P_1 and Q, with an additional constraint that tunes the new equilibrium. The solution P^* is analog to the escort distribution, that appears quite arbitrary [4, 5] in non-extensive statistics.
Then I present two scenarii for the mean observation constraint: the observable is taken as the mean under P_1, the distribution of a "subsystem", or under P^*, the apparent distribution of the system. In the two cases, I show that the amended MaxEnt formulation leads to the maximization of the Rényi α Q-entropy, subject to the corresponding mean constraint. So doing, we recover two of the classical choices of constraint in the nonextensive literature. These two scenarii lead to two Tsallis-like distributions with opposite exponents, and the entropic index α appears to be simply the Lagrange parameter associated to the log-likelihood constraint.
A difficulty comes from the determination of the parameter of the Tsallis-like distributions, that are self-referential. In order to identify the value of their natural parameter, I propose two 'alternate' (but effectively computable) dual functions, whose maximizations enable to exhibit the optimum parameters.
In the conference, I will illustrate these results for several references $Q(x)$ present the results of numerical evaluations, and recover some well-known entropy functionals in the classical Gibbs-Boltzmann-Shannon limit. We will also give further results concerning symmetry and duality between the solutions associated with classical and generalized mean constraint, and between entropy functionals. Finally I will discuss the underlying Legendre structure of generalized thermodynamics associated to this setting.

References

[1] C. Tsallis, "Nonextensive statistics: Theoretical, experimental and computational evidences and connections," *Brazilian Journal of Physics*, vol. 29, pp. 1–35, March 1999.

[2] C. Tsallis, "Entropic nonextensivity: a possible measure of complexity," *Chaos, Solitons,& Fractals*, vol. 13, pp. 371–391, 2002.

[3] C. Tsallis, "Possible generalization of Boltzmann-Gibbs statistics," *Journal of Statistical Physics*, vol. 52, no. 1-2, pp. 479–487, 1988.

[4] B. R. La Cour and W. C. Schieve, "Tsallis maximum entropy principle and the law of large numbers," *Phys. Rev. E*, vol. 62, pp. 7494 – 7496, November 2000.

[5] J. Grendar, M. and M. Grendar, "Maximum entropy method with non-linear moment constraints: challenges," in *Bayesian inference and maximum entropy methods in science and engineering* (G. Erickson, ed.), AIP, 2004.

[6] C. Beck, "Generalized statistical mechanics of cosmic rays," *Physica A*, vol. 331, pp. 173–181, january 2004.

[7] G. Wilk and Z. Wodarczyk, "Interpretation of the nonextensitivity parameter q in some applications of Tsallis statistics and Lévy distributions," *Physical Review Letters*, vol. 84, pp. 2770–2773, March 2000.

NONPARAMETRIC BAYESIAN ESTIMATION OF X/γ-RAY SPECTRA USING A HIERARCHICAL POLYA TREE – DIRICHLET MIXTURE MODEL

Éric Barat[1], Thomas Dautremer[1]
(1) DeTeCS, CEA-Saclay, France

Abstract

We address the problem of X/γ-ray spectra estimation in the fields of nuclear physics and astrophysics. Bayesian estimation of experimental backgrounds has been studied in [1] involving splines. Since Dirichlet Processes (DP) sit on discrete measures, they provide an appealing prior for photopeaks. On the other hand, in order to tackle the complexity of experimental backgrounds, we consider a Polya Tree Mixture (PTM) – with suitable parameters yielding distribution continuity – for which predictive densities exhibit better smoothness properties than a single Polya Tree. Furthermore, it is easy to introduce some physical Compton line approximation formula (e.g. Klein-Nishina) in the base measure of the Polya Tree, or some physically driven local modifications of the PTM prior parameters. As backgrounds depend on photopeaks locations, we propose a hierarchical model where the PTM is conditioned on the DP. We use a beta prior for the mixing proportion between the DP and the PTM. Energies are not directly observed due to detection devices noises which introduce a convolution of both discrete and continuous measures by an assumed gaussian kernel whose variance is an unknown linear function of energy. Thus, the proposed semiparametric model for experimental data becomes a hierarchical Polya Tree–Dirichlet mixture of normal kernels. The quantities of interest for physicists are usually posterior functionals of the DP mixing distribution. This implies an inverse problem which is carried out in the framework of finite stick-breaking representation. To allow finite representation of PTM, we assume infinite prior parameters after a certain stage. The *blocked Gibbs sampler* of [2] is extended to update simultaneously the hidden allocation variables either from a DP component or from a set of the PTM latest informative partition. Thanks to conjugacy and conditioning on the hidden allocation variables, draws from the posterior DP and PTM are easily obtained. With minor modifications the algorithm can deal with binned data which turns out to be computationally attractive for huge datasets.

References:

[1] R. Fischer, K. Hanson, V. Dose and W. von der Linden, *Physical Review E* **61**, 1152–1160 (2000).

[2] H. Ishwaran and L. James, *Journal of the American Statistical Association* **96**, 161–173 (2001).

Key Words: Bayesian nonparametrics, Polya tree, Dirichlet mixture, X/γ spectra.

NONPARAMETRIC BAYESIAN ESTIMATION OF CENSORED COUNTERS INTENSITY FROM THE INDICATOR DATA

T. Dautremer[1], É. Barat[1], T. Trigano[2]
(1) DeTeCS, CEA Saclay, France
(2) Department of statistics, Hebrew University of Jerusalem, Israel

Abstract

Physical counting devices are usually imperfect in the sense that they are unable to record all particles present at their input. After a particle is registered, the counter is inhibited for a positive duration. This censoring period is referred to as dead time (or busy period). Following [1], we denote by $N(t)$ the number of recorded particles at t and $M(t)$ the input poisson process with intensity function λ_t, then $N(t) = \int_0^t Y(s)\, dM(s)$ where $Y(t) = I(t \geq S_{N(t-)} + C_{N(t-)})$, S_i is the time of the i^{th} recorded particle and C_i the corresponding dead time. We thus consider the problem of estimating λ_t given a sample path of the indicator data Y. Since it might be cumbersome to infer from the busy distribution, we propose a bayesian nonparametric method leaning only on idle periods. For all t where $Y(t) = 0$, we define the *lifetime* $t^\star = t - \sum_{j=1}^{N(t-)} C_j$. A Polya tree [2] prior defined over the lifetime space is used for the normalized intensity. Due to the data-dependent partition, the problem leads to a nonhomogeneous poisson intensity $\lambda^\star_{t^\star}$ estimation. With a gamma prior for the integrated intensity, the posterior remains the product of a Polya tree and a gamma distribution. The intensity for the idle periods is then achieved by setting $\lambda_t = \lambda^\star_{t^\star}$. For busy periods, an interpolation scheme can be used. For application purposes involving an open-ended stream, we propose an estimator of the intensity based on the posterior expectation of a shifted polya trees finite mixture which leads to a finite response nonlinear filter. Assuming small λ_t variations during the busy periods, this method is suitable for various kinds of censoring mechanism because no assumption is made about the dead time distribution.

For the usual case of type-II counter [3], formally an $M_t/G/\infty$ queue, $Y(t) = I(Q(t) > 0)$ with $Q(t)$ the number of clients in the queue, we improve the method by using additional information from the distribution of the number of particles participating to a busy period conditionally to its length.

References:

[1] Y. Kim, Annals of statistics **27**, 562-588 (1999).
[2] M. Lavine, Annals of statistics **20**, 1222-1235 (1992).
[3] R. Pyke, Annals of Mathematical Statistics **29**, 737-754 (1958).

Key Words: Bayesian nonparametrics, Type-II counter, $M_t/G/\infty$, Polya tree.

MaxEnt Velocity Profiles in Laminar to Turbulent Flow

Robert K. Niven
School of Aerospace, Civil and Mechanical Engineering
The University of New South Wales at ADFA
Northcott Drive, Canberra, ACT, 2600, Australia.
Email: r.niven@adfa.edu.au
28 February 2006

Abstract

This work applies the differential equation method developed by Chiu and co-workers[1] - based on Jaynes' maximum entropy method[2] - to determine the "most probable" steady-state velocity profile $u(y)$ in three systems of "classical" fluid mechanics: (i) axial flow in a cylindrical pipe (Poiseuille flow) (previously examined by Chiu[1]); (ii) flow between stationary parallel plates; and (iii) flow between moving parallel plates (Couette flow). In each case, the analysis yields an analytical solution for the velocity profile over the complete spectrum of laminar to turbulent flow. The profiles are obtained as functions of the maximum velocity u_m and parameter $M = -u_m \lambda_1$, where λ_1 is the Lagrangian multiplier for the conservation of mass constraint. M can be interpreted as a "temperature of turbulence", with M=0 indicating laminar flow and $M \to \infty$ complete turbulence. The main elements of this analysis, which have been presented briefly[3], are reproduced here.

For the axial flow system, the predicted profiles and their moments reduce to the well-known laminar solution at M=0. For M>0, the resulting solution can be used in place of existing semi-empirical correlations for the velocity profile in axial flow[1,4]. For the plane parallel flows, in order to match both the laminar profiles and higher order moments at M=0, it is necessary to make use of the relative entropy (Kullback-Liebler cross-entropy) function, incorporating a different Bayesian prior (Jaynes' invariant) distribution. A method to determine this prior distribution is described.

The analysis is then used to derive a new maximum-entropy laminar-turbulent boundary layer theory, for the velocity profile in steady flow along a flat plate. For M=0, this reduces to the laminar boundary layer theory given in some texts[4], which approximates the Prandtl-Blasius solution to the Navier-Stokes equation[5]. For turbulent flow, it yields a previously unreported solution.

Keywords: MaxEnt; fluid mechanics; velocity profile; turbulent flow; boundary layers.

References:

1 C.-L. Chiu (1987) *J. Hydraul. Eng.-ASCE* 113(5) 583; C.-L. Chiu (1988) *J. Hydraul. Eng.-ASCE* 114(7): 738; C.-L. Chiu, G.-F. Lin, J.-M. Lu, (1993) *J. Hydraul. Eng.-ASCE* 119(6): 742.
2 E.T. Jaynes (1957) *Phys. Rev.* 106: 620.
3 R.K. Niven (2005) 3rd Int. Conf.: NEXT-Sigma-Phi, 3-18 August 2005, Kolymbari, Crete, Greece.
4 R.L. Street, G.Z. Watters, J.K. Vennard (1996) Elementary Fluid Mechanics, 7th ed., John Wiley, NY.
5 H. Schlichting, K. Gersten (2001), Boundary Layer Theory, 8th ed., Springer, NY.

Bayesian Methods in Ground Penetrating Radar

Louis Roemer and David Cowling

Abstract: A low frequency interferometer is used to collect data on subsurface obstacles. A simple model for the near-field reflection allows computation of the probability of the target location. The antennas of the interferometer are orthogonally polarized, and a balancing mechanism allows minimizing direct transmission of signals. Thus, the interferometer shows, mainly, reflections from changes in the geometry.

Applications tested were land mine location and utility pipe location. An added benefit is that reflections from layers of soil, due to the constant reflection from the layer, do not appear as distracting false targets.

Scheduling of schools

Tj. Romke Bontekoe[1], Do Kester[2], John Skilling[3]
(1) Bontekoe Research, Rooseveltstraat 4-d, 2321 BM Leiden, The Netherlands
(2) SRON, Postbus 800, 9700 AV Groningen, The Netherlands
(3) Killaha East, Kenmare, Kerry, Ireland
(e-mail: romke@bontekoe.nl)

Abstract

The scheduling of schools where students, teachers, rooms and their lessons according the curriculum is a large combinatorial optimization problem. It appears that there *many* solutions for a medium sized Dutch secondary school (1200 students, 100 teachers, 50 rooms). Therefore one can optimize, i.e. select a better solution from the lot. We have developed a computational method to find such solutions.

We distinguish between "hard wishes" and "soft wishes". In order to have a valid schedule all hard wishes must be fulfilled, such as the lessons table, student clusterings, teacher availability, room restrictions, etc. Even if a only single lesson cannot be placed, the schedule is invalid.

In fact, very many valid school schedules are computed and as many "soft wishes" as possible are fulfilled. The main soft wish is minimization of idle hours for students. There are many more soft wishes for which the relative importance can be adjusted. We find solutions which comply with all hard wishes and a balanced compromise between the soft wishes.

Key Words: Combinatorics, Optimization, School, Schedule

Electromagnetic Induction Landmine Detection using Bayesian Model Comparison

Paul M. Goggans and Ying Chi
The University of Mississippi
Electrical Engineering Department
University, MS 38677

Electromagnetic induction (EMI) landmine detection can be cast as a Bayesian model comparison problem. The models used for low metallic-content mine detection are based on the equivalent electrical circuit representation of the EMI detection system. The EMI detection system is characterized and modeled by the impulse response of its equivalent circuit. The analytically derived transfer function between the transmitter coil and receiver coil demonstrates that the EMI detection system is a third order system in the absence of a mine and that the presence of a mine adds an additional pole that makes the detection system fourth order. The value of the additional pole is determined by the equivalent inductance and resistance of the mine and is unique for each mine. This change in system order suggests that measured system impulse responses can be used in conjunction with impulse response models to infer the presence or absence of a landmine. The difficulty of this techniques is that the amplitude of the term added to the the system impulse response by the landmine is small compared to the impulse response of the system alone. To test the feasibility of Bayesian inference based EMI landmine detection, an EMI detection system experiment was designed and built. In the experiment the EMI detection system was driven by a broadband maximal-length sequence (MLS) in order to obtain sufficient dynamic range in the measured impulse responses. This paper discusses the development of parameterized impulse response models for the detections system with and without a landmine present and the assignment of appropriate priors for the parameters of these models. This paper also presents the ratios of computed posterior probabilities for the mine and no mine models based on data obtained from the experimental EMI landmine detection system. These odds ratios demonstrate the potential of Bayesian EMI landmine detection.

Online Learning in Discrete Hidden Markov Models

Roberto C. Alamino[1], Nestor Caticha[2]
(1) Aston University, Birmingham, UK
(2) University of Sao Paulo, Sao Paulo, Brazil

Abstract

We present and analyze three different online algorithms for learning in discrete Hidden Markov Models (HMMs) and compare their performance with the Baldi-Chauvin Algorithm. Using the Kullback-Leibler divergence as a measure of the generalization error we draw learning curves in simplified situations and compare the results. The performance for learning drift concepts of one of the presented algorithms is analyzed and compared with the Baldi-Chauvin algorithm in the same situations. A brief discussion about learning and symmetry breaking based on our results is also presented. Key Words: HMM, Online Algorithm, Generalization Error, Bayesian Algorithm.

A BAYESIAN APPROACH TO CALCULATING FREE ENERGIES OF CHEMICAL AND BIOLOGICAL SYSTEMS

Andrew Pohorille
NASA-Ames Research Center, USA
(e-mail: pohorill@max.arc.nasa.gov, fax: 650-604-1088)

Abstract

A common objective of molecular simulations in chemistry and biology is to calculate the free energy difference, ΔA, between states of a system of interest. Important examples are protein-drug interactions, protein folding and ionization states of chemical groups. However, accurate determination of ΔA from simulations is not simple. This can be seen by representing ΔA in terms of a one-dimensional integral of $exp(-\Delta E/k_BT) \times P(\Delta E)$ over ΔE. In this expression, ΔE is the energy difference between two states of the system, $P(\Delta E)$ is the probability distribution of ΔE, k_B is the Boltzmann constant and T is temperature. For finite systems, $P(\Delta E)$ is a distorted Gaussian. Note that the exponential factor weights heavily the low ΔE tail of $P(\Delta E)$, which is usually known with low statistical precision.

One way to improve estimates of ΔA is to model $P(\Delta E)$. Generally, this approach is rarely successful. Here, however, we take advantage of the "Gaussian-like" shape of $P(\Delta E)$. As is known in physics, such a function can be conveniently represented by the square of a "wave function" which is a linear combination of Gram-Charlier polynomials. The number of terms, N, in this expansion supported by the data must be determined separately. This is done by calculating the posterior probability, $P(N/\mathbf{\Delta E})$, where $\mathbf{\Delta E}$ stands for all sampled values of ΔE. In brief, the dependence of the likelihood function on the coefficients of the expansion, $\mathbf{C_N}$ is marginalized by determining their optimal values using Lagrange multipliers, and then expanding $P(\mathbf{\Delta E})/\mathbf{C_N}, \mathbf{N})$ around the optimal solution. Special care needs to be taken to ensure convergence of this expansion. As expected, the maximum likelihood solution consists of two terms. One is related to the optimal values of $\mathbf{C_N}$ and always increases with N. The second term is an "Ockham's Razor" penalty. It involves a multivariate Gaussian integral on the N-dimensional hypersphere, which arises due to mormalization. This integral cannot be calculated analytically, but accurate approximations, which properly account for problem symmetries, can be obtained.

The method offers the largest improvements over conventional approaches when $P(\Delta E)$ is broad and sample size is relatively small. This makes is particularly suitable for computer aided drug design, in which the goal is to screen rapidly a large number of potential drugs for binding with the protein target.

EXPLORING THE CONNECTION BETWEEN SAMPLING PROBLEMS IN BAYESIAN INFERENCE AND STATISTICAL MECHANICS

Andrew Pohorille
NASA-Ames Research Center, USA
(e-mail: pohorill@max.arc.nasa.gov, fax: 650-604-1088)

Abstract

The Bayesian and statistical mechanical communities often share the same objective in their work – estimating and integrating probability distribution functions (pdfs) describing stochastic systems, models or processes. Frequently, these pdfs are complex functions of random variables exhibiting multiple, well separated local minima. Conventional strategies for sampling such pdfs are inefficient, sometimes leading to an apparent non-ergodic behavior. Several recently developed techniques for hadling this problem have been successfully applied in statistical mechanics.

In the multicanonical and Wang-Landau Monte Carlo (MC) methods, the correct pdfs are recovered from uniform sampling of the parameter space by iteratively establishing proper weighting factors connecting these distributions. Trivial generalizations allow for sampling from any chosen pdf. The closely related transition matrix method relies on estimating transition probabilities between different states. All these methods proved to generate estimates of pdfs with high statistical accuracy. In another MC technique, parallel tempering, several random walks, each corresponding to a different value of a parameter (e.g. "temperature"), are generated and occasionally exchanged using the Metropolis criterion. This method can be considered as a statistically correct version of simulated anneling.

An alternative approach is to represent the set of independent variables as a Hamiltonian system. Considerable progress has been made in understanding how to ensure that the system obeys the equipartition theorem or, equivalently, that coupling between the variables is correctly described. Then a host of techniques developed for dynamical systems can be used. Among them, probably the most powerful is the Adaptive Biasing Force method, in which thermodynamic integration and biased sampling are combined to yield very efficient estimates of pdfs.

The third class of methods deals with transitions between states described by rate constants. These problems are isomorphic with chemical kinetics problems. Recently, several efficient techniques for this purpose have been developed based on the approach originally proposed by Gillespie.

Although the utility of the techniques mentioned above for Bayesian problems has not been determined, further research along these lines is warranted.

Inverse Problem For Estimating Heat Source

A. Neisy

Department of Mathematics and Statistics , Faculty of Economics, Allameh .Tabatabaie University. Dr. Beheshti Ave., Tehran, Iran

Abstract: This paper considers, a two-dimensional inverse heat conduction problem. The direct problem will be solved by an application of the heat fundamental solution, and the heat source to be estimated by using least-square .method

Key Words: two-dimensional problem, direct and inverse heat conduction .problem, overposed data

Phase space methods in continuous tensor products of Hilbert spaces

A. Vourdas
Department of Computing,
University of Bradford,
Bradford BD7 1DP, United Kingdom

A continuum of quantum mechanical oscillators in a straight line, is considered. The Hilbert space of the system is a continuous tensor product of the Hilbert spaces describing the individual oscillators.

A mode-position operator u_x is introduced whose expectation value describes the location of a quantum state within the chain of oscillators. A mode-momentum operator u_p is also introduced whose expectation value describes the change of the mode position with time. Both of these operators involve all oscillators and are very different from the position and momentum operators of individual oscillators. Their exponentials are displacement operators which propagate a quantum state in the continuum of oscillators; and also change its momentum. They are collective transformations and are very different from displacements in the phase space of individual oscillators.

Entropic quantities which describe correlations and entanglement between the various oscillators are also studied.

The work is presented in a quantum mechanical language; but it is also applicable to related areas like time-frequency analysis in signal processing, applied harmonic analysis, etc. In this more general context it studies phase-space methods for problems described with a continuous tensor product of the Hilbert spaces.

Title: Data, Virtual Data, and Anti-Data.

Author: Carlos C. Rodriguez

Abstract

\delta-priors optimize natural notions of ignorance. When the likelihood is in the exponential family the 0-priors become the standard cojugate priors relative to the information volume. In this case prior information is equivalent to having \alpha > 0 extra virtual observations. On the other hand 1-priors are not conjugate and where the 0-priors add the \alpha virtual observations to the actual n sample points, the 1-priors subtract the \alpha from the n. I call this "anti-data" since \alpha of these points annihilate \alpha of the observations leaving us with a total of n-\alpha. Thus, 1-priors are more ignorant than 0-priors. True ignorance, that claims only the model and the observed data, has a price. To build the prior we must spend some of the information cash in hand. No free lunches.

ENTROPY AND SEMI-MARKOV PROCESSES

Valerie Girardin,
LMNO Université de Caen, BP5186, 14032 Caen, France
(e-mail: girardin@math.unicaen.fr)

Abstract

Entropy and Markov processes are linked since the first version of the asymptotic equirepartition property (AEP) stated by Shannon in 1948 for Markov chains. We define explicitely the entropy rate for semi-Markov processes and extend the AEP or ergodic theorem of information theory to these nonstationary processes.

Among a given collection of functions satisfying constraints, selecting the one with the maximum entropy is equivalent to adding the less of information possible to the considered problem. The definition of an explicit entropy rate for processes allows one to extend the maximum entropy method to this case. We study different problems for Markov and semi-Markov processes, illustrated in reliability, queueing theory, sismology...

References:

[1] V. Girardin (2004) ENTROPY MAXIMIZATION FOR MARKOV AND SEMI-MARKOV PROCESSES *Methodology and Computing in Applied Probability*, V6, 109–127.

[3] V. Girardin (2005) ON THE DIFFERENT EXTENSIONS OF THE ERGODIC THEOREM OF INFORMATION THEORY *Recent Advances in Applied Probability*, R. Baeza-Yates, J. Glaz, H. Gzyl, J. Hüsler & J. L. Palacios (Eds), Springer-Verlag, San Francisco, pp163–179.

[3] V. Girardin & N. Limnios (2006). ENTROPY FOR SEMI-MARKOV PROCESSES WITH BOREL STATE SPACES: ASYMPTOTIC EQUIREPARTITION PROPERTIES AND INVARIANCE PRINCIPLES *Bernoulli*, to appear.

Key Words: asymptotic equirepartition property, entropy rate, Markov chains, Markov processes, maximum of entropy, semi-Markov processes, Shannon-McMillan-Breiman theorems.

ENTROPY COMPUTATION IN PARTIALLY OBSERVED MARKOV CHAINS

François Desbouvries
Institut National des Télécommunications, Evry, France
e-mail: Francois.Desbouvries@int-evry.fr

Abstract

Hidden Markov Chains (HMC) [1] are widely used in speech recognition, image processing or protein sequence analysis, due to early availability of efficient Bayesian restoration (Forward-Backward, Viterbi) or parameter estimation (Baum Welch) algorithms. More recently, the problem of computing in an HMC the entropy of the possible hidden state sequences that may have produced a given sequence of observations has been addressed, and an efficient (i.e., linear in the number of observations) algorithm has been proposed [2].

Among possible extensions of HMC, Pairwise (PMC) [3] and Triplet [4] Markov Chains (TMC) have been introduced recently. In a TMC we assume that $t = (x, r, y)$, where x is the hidden process, y the observation and r a latent process, is a Markov chain (MC). So a TMC can be seen as a vector MC, in which one observes some component y and one wants to restore some part of the remaining components. In a TMC the marginal process (x, r) is not necessarily an MC, but the conditional law of (x, r) given the observations y is an MC; as in HMC, this key computational property enables the development of efficient restoration or parameter estimation algorithms. In this paper, we extend to TMC the entropy computation algorithm of [2]. The resulting algorithm remains linear in the number of observations.

References:

[1] Y. Ephraim and N. Merhav, "Hidden Markov processes", *IEEE Transactions on Information Theory*, vol. 48-6, pp. 1518-69, June 2002.

[2] D. Hernando, V. Crespi and G. Cybenko, "Efficient Computation of the Hidden Markov Model Entropy for a Given Observation Sequence", *IEEE tr. Info. Th.*, pp. 2681-85, July 2005

[3] W. Pieczynski, "Pairwise Markov Chains", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 25-5, pp. 634-39, May 2003

[4] W. Pieczynski and F. Desbouvries, "On Triplet Markov chains", *Proceedings of the International Symposium on Applied Stochastic Models and Data Analysis (ASMDA 2005)*, Brest, France, May 17-20, 2005

Key Words: Entropy, Hidden Markov Chains, Markovian models

BAYESIAN SMOOTHING ALGORITHMS IN PARTIALLY OBSERVED MARKOV CHAINS

Boujemaa Ait-el-Fquih and François Desbouvries
Institut National des Télécommunications, Evry, France
e-mail: Francois.Desbouvries@int-evry.fr

Abstract

An important problem in signal processing consists in estimating an unobservable process x from an observed process y. In Hidden Markov Chains (HMC), efficient Bayesian smoothing restoration algorithms have been proposed in the discrete [1] as well as in the Gaussian case [2] [3].

Among other extensions of HMC, Triplet Markov Chains (TMC) have been introduced recently (see e.g. [4]). In a TMC we assume that the triplet (x, r, y) (in which r is some additional process) is a Markov Chain (MC). So a TMC can be seen as a vector MC, in which one observes some components y and one wants to restore some part of the remaining components. In a TMC the marginal process (x, r) is not necessarily an MC, but the conditional law of (x, r) given the observations y is an MC; as in HMC, this key computational property enables the development of efficient restoration or parameter estimation algorithms. This paper addresses fixed-interval smoothing algorithms in TMC and is a continuation of the work of [5]. In particular, we extend to Gaussian TMC the Bryson and Frazier algorithm, the backward-forward RTS algorithm, the Fraser and Potter algorithm and the backward-forward RTS algorithm of Desai *et al.*

References:

[1] L. R. Rabiner, "A Tutorial on Hidden Markov Models and Selected Applications in Speech Recognition, *Proceedings of the IEEE*, February 1989.

[2] T. Kailath, A. H. Sayed and B. Hassibi, *Linear estimation*, Prentice Hall, Upper Saddle River, New Jersey 2000

[3] H. L. Weinert, *Fixed interval smoothing for state-space models*, Kluwer Academic Publishers, Boston 2001.

[4] W. Pieczynski and F. Desbouvries, On Triplet Markov chains, *Proceedings of the International Symposium on Applied Stochastic Models and Data Analysis (ASMDA 2005)*, Brest, France, May 17-20, 2005

[5] B. Ait-el-Fquih and F. Desbouvries, "Bayesian smoothing algorithms in Pairwise and Triplet Markov chains", *Proceedings of the 2005 IEEE Workshop on Statistical Signal Processing*, Bordeaux, France, July 2005.

Key Words: Hidden Markov Chains, state-space models, Markovian models, smoothing algorithms.

Title: Applications of Maximum Entropy in Gravitational Wave Astronomy

Author: Lee Samuel Finn
Institution:
Center for Gravitational Wave Physics,
The Pennsylvania State University

Abstract:
Large gravitational wave detectors in the United States (LIGO) and Europe (GEO and Virgo) have just now reached a sensitivity that makes them sensitive to gravitational wave emission from astronomical phenomena. In the next decade, ESA and NASA plan to place an even more sensitive detector (LISA) into space. Sources that these detectors have the capability of observing include the formation of the black holes that are thought to power gamma-ray bursts, the stellar core collapses that power type II supernovae, the coalescence of supermassive black holes that follows the collision of their host galaxies, and the myriad of compact binary white dwarf binary systems that populate our galaxy. The first detection of gravitational waves by these detectors will usher in the era of gravitational wave astronomy: the use of gravitational waves as a tool of astronomical discovery.

Gravitational wave detectors are not imaging instruments and individual gravitational wave detectors lack the ability to localize a source on the sky. From a network of detectors we can synthesize a beam and thus determine the position of a source and the radiation amplitude and phase in each gravitational wave polarization. Alternatively, the signal acquired from a detector that is moving with respect to a source will be phase and amplitude modulated in a manner that depends on the source's sky location, the signal in each polarization, and the detector's changing position and orientation with respect to the source. From this information and models of the radiation expected from different sources we can test general relativity and learn about the sources we are observing. Here we describe the development and use of maximum entropy based tools to recover the gravitational wave signal amplitude and phase in each polarization, and the sky location of one or more sources whose radiation is incident on an array of detectors or a moving detector. These tools are being used now for the analysis of data from the United States LIGO detector and will likely play an important role in the analysis of data from the LISA detector.

Gibbs Paradox and the Higher Similarity - Higher Entropy Relationship

Shu-Kun Lin
Molecular Diversity Preservation International (MDPI),
Matthaeusstrasse 11, CH-4057 Basel, Switzerland
(e-mail: lin@mdpi.org, http://www.mdpi.org/lin)

Abstract

There are three kinds of correlation of the entropy of mixing with similarity. The Gibbs paradox statement, which has been regarded as a very fundamental assumption in statistical mechanics, says that the entropy of mixing or assembling to form solid assemblages, liquid and gas mixtures or any other analogous assemblages such as quantum states, decreases discontinuously with the increase in the property similarity of the composing individuals. Most authors accept this relastionship (e.g. [1]). Some authors revised the Gibbs paradox statement and argued that the entropy of mixing decreases continuously with the increase in the property similarity of the individual components [2]. A higher similarity - higher entropy relationship and a new theory has been constructed: entropy of mixing or assembling increases continuously with the increase in the similarity. Similarity Z can be easily understood when two items A and B are compared: if A and B are distinguishable (minimal similarity), Z=0. If they are indistinguishable (maximal similarity), Z=1.

References:

[1] E. T. Jaynes, The Gibbs Paradox, In Maximum Entropy and Bayesian Methods; C. R. Smith, G. J. Erickson, P. O. Neudorfer, Eds.; Kluwer Academic: Dordrecht, 1992, p.1-22.

[2] J. von. Neumann, Mathematical Foundations of Quantum Mechanics. Princeton, NJ: Princeton University Press, 1955.

Key Words: Gibbs Paradox, Entropy, configurational entropy

AN APPLICATION OF ENTROPIC DYNAMICS ON CURVED STATISTICAL MANIFOLDS

Carlo Cafaro, S. A. Ali, A. Giffin
Department of Physics, University at Albany-SUNY,
1400 Washington Avenue, Albany, NY 12222, USA
(e-mail: carlocafaro2000@yahoo.it)

Abstract

Any attempt to unify the classical theory of gravity with quantum theories of electromagnetic, weak and strong forces in a single unified theory has been unsuccessful so far. Entropic Dynamics (ED), namely the combination of principles of inductive inference (Maximum Entropy Methods) and methods of Information Geometry (IG), is a theoretical framework constructed to explore the possibility that laws of physics, either classical or quantum, might be laws of inference rather than laws of nature. The ultimate goal of such an ED concerns the derivation of Einstein's theory of gravity from an underlying "statistical geometrodynamics" [1].

Our objective here is to show explicitly all the steps needed to derive an ED model and to underline the most delicate aspects of it. The first step is to identify the appropriate variables describing the system, and thus the corresponding space of macrostates. This is by far the most difficult step because there does not exist any systematic way to search for the right macro variables; it is a matter of taste and intuition, trial and error. In the ED model here presented we do not specify the nature of our system, it might be a thermal system or something else. We will make connections to conventional physical systems only later in the formulation of the ED model. We only assume that the space of microstates is 2D and that all the relevant information to study the dynamical evolution of such a system is contained in a 3D space of macrostates. The second step is to define a quantitative measure of change from one macrostate to another. Maximum Entropy Methods lead to the assignment of a probability distribution to each macrostate, while methods of IG lead to the assignment of the Fisher-Rao information metric quantifying the extent to which one distribution can be distinguished from another. The ED is defined on the space of probability distributions $\mathcal{M}_s$. The geometric structure of $\mathcal{M}_s$ is studied in detail. We show that $\mathcal{M}_s$ is a 3D pseudosphere with constant negative Ricci scalar curvature, $R = -1$. The final step concerns the study of irreversible and reversible aspects of such ED on $\mathcal{M}_s$. In the former case, we study the evolution of the system from a given macrostate to an unknown final macrostate. This study is used to show that the microstates of the model undergo an irreversible diffusion process. In the latter, we study the evolution of the system from a given initial macrostate to a given final state. The trajectories of the system are shown to be hyperbolic curves on $\mathcal{M}_s$, and the surface of evolution of the statistical parameters describing $\mathcal{M}_s$ is plotted. Finally, similarities and possible connections between ED methods and established physics are highlighted.

References:

[1] A. Caticha: "The Information Geometry of Space and Time", Presented at MaxEnt2005, the 25th International Workshop on Bayesian Inference and Maximum Entropy Methods (August 7-12, 2005, San Jose, California, USA).

Reconstruction of the Electron Energy Distribution Function from Optical Emission Spectroscopic Measurements

Dirk Dodt[1], Andreas Dinklage[1], Rainer Fischer[2]
(1) Institut für Plasma Physik, Teilinstitut Greifswald
(2) Institut für Plasma Physik, Teilinstitut Garching
(e-mail: ddodt@ipp.mpg.de)

Abstract

The properties of a low temperature plasma (as for example used in energy saving light bulbs) are mainly determined by the energy distribution of the free electrons. This distribution is described by the so-called electron energy distribution function (EEDF). A well established method to obtain the EEDF is to measure the current-voltage characteristics of a plasma using a small wire in contact with the plasma (probe). The approach presented here is motivated by the idea to utilise the light emitted by excited gas atoms, in order to get rid of the perturbing probe brought into the plasma.

The inference of the EEDF from the measured intensities is an example of an ill-posed inversion problem, because of the high sensitivity of the reconstruction on small errors of the line intensities.

The forward calculation consists of a so-called stationary collisional-radiative model which is describing the interaction of atoms and ions with the free electrons and the discharge device.

The systematic uncertainties in the model parameters, namely the different atomic data that enter the calculation, have to be considered with particular care.

First results are shown for the spectrum of a neon discharge lamp. The radially averaged EEDF is reconstructed. The applicability of different functional forms of the EEDF is assessed. In a first step Maxwell and Druyvenstein distributions which are having only a small number of parameters are considered.

Key Words: Plasma Physics, Applied Bayesian Data Analysis

OPTIMIZATION OF PLASMA DIAGNOSTICS USING BAYESIAN PROBABILITY THEORY

H. Dreier[1], A. Dinklage[1], R. Fischer[2], M. Hirsch[1], P. Kornejew
Max–Planck–Institut für Plasmaphysik, EURATOM Association,
[1] Teilinstitut Greifswald, D–17491 Greifswald, Germany
[2] D–85748 Garching, Germany
(e-mail: hdreier@ipp.mpg.de)

Abstract

The Wendelstein 7-X stellarator will be a magnetic fusion device and is presently under construction. Its diagnostic set-up is currently in the design process to optimize the outcome under given technical constraints. In general, the preparation of diagnostics for magnetic fusion devices requires a physical model of the measurement which relates the physical effect to the measured data (*forward function*), and a diagnostics model which describes the error statistics.

The approach presented here bases on maximization of an information measure (Kullback-Leibler entropy, see ref. [1]). Bayesian probability theory allows one to link measures from information theory with the model for the fusion diagnostics. The approach can be considered as the implementation of a *virtual diagnostic* which generates data from a range of parameters. The virtual diagnostic employs the forward function and accounts for the error statistics. Then, optimization means maximization of the expected utility with respect to the design parameters. It allows for extensive design studies of effects due to physical input and possible benefits due to technical elements. Comparisons with other information measures and approximation methods for the prior predictive value are discussed.

The reconstruction of density profiles by means of a multichannel infrared interferometer at W7-X is investigated in detail. The influence of different error statistics and the robustness of the result are discussed. In addition, the impact of technical boundary conditions is shown.

References:

[1] R.Fischer, AIP Conf. Proc. 735 (2004) 76-83

[2] H.Dreier et al., Fusion Science and Technologies, (accepted 2006)

Key Words: diagnostic design, optimization, information measure

On the theory of phase transition Landau

Gadjiev B.R.

International University of Nature, Society and Man,
141980, 19 Universitetskaya str., Dubna, Moscow Region, Russia
gadjiev@uni-dubna.ru

We consider close-packed structure with defects, which undergoes structural phase transition, and we research dependence of critical exponents on concrete kind of defects distribution. The analysis shows, that the connectivity distribution function of defects in structure can be presented by the generalized Boltzmann factor, as it is done in superstatistics. Namely, the distribution of defects in close-packed structure can be considered as the process of homogeneous growth. Using a maximum entropy principle it is possible to show, that in this case we have an exponential distribution of defects connectivity. If the spatial distribution of defects in structure is random, then, generally, the number of entering links will be random variable. For example, if this distribution is gamma distribution we obtain the analogue of Tsallis distributions. Other spatial distributions of defects generate infinite number different distributions. For a statistical mechanical foundation we use a maximum entropy principle, which allows to obtain the generalized Boltzmann factor that allows to obtain the concrete distribution links of defects. Namely from a entropy functional on which the constraints we are imposed and we define the distribution function of defects in the system. The maximum entropy principle, which was used in no extensive statistical mechanics, allows derive the generalized Boltzmann factor. It allows to obtain various distribution functions connectivity of defects in structure by a unified way.

For the analysis of defects distribution dependence critical exponents we introduce a free energy functional, which depends on an order parameter and on connectivity distribution of defects of structure. The symmetry of an order parameter is defined by an irreducible representation of a space group of structure. After the procedure an average of a free energy functional on connectivity of defects (practically it implies calculation a moments of the distribution) we obtain, that the critical behaviour strongly depends on the form of the distribution function and can essentially differ from of the mean-field behaviour. The analysis of experimental results shows, that such situation is characteristic for doped layered crystals (system with competing interaction).

MARGINALIZED MAXIMUM A POSTERIORI HYPER-PARAMETER ESTIMATION FOR GLOBAL OPTICAL FLOW TECHNIQUES

Kai Krajsek, Rudolf Mester
Institute for Computer Science,D-60054 Frankfurt, Germany
(e-mail: krajsek@vsi.cs.uni-frankfurt.de
http://www.vsi.cs.uni-frankfurt.de)

Abstract

Global optical flow estimation methods contain a regularization parameter (or prior and likelihood hyper-parameters if we consider the statistical point of view) which control the tradeoff between the different constraints on the optical flow field. Although experiments (see e.g. Ng et al. [2]) indicate the importance of the optimal choice of the hyper-parameters, only little attention has been focused on the optimal choice of these parameters in global motion estimation techniques in literature so far (the authors are only aware of one contribution [2] which attempts to estimate only the prior hyper-parameter whereas the likelihood hyper-parameter needs to be known). We adapted the marginalized maximum a posteriori (MMAP) estimator developed in [1] to simultaneously estimating hyper-parameters and optical flow for global motion estimation techniques. The optimal hyper-parameters are strongly determined by first order statistics in the image scene, i.e. the illumination distribution. Optimal values for the hyper-parameter of former image scenes could therefore be used to feed in the Bayesian hyper-parameter estimation framework. Furthermore, the resulting objective function is not convex with respect to the hyper-parameters, thus an appropriate starting point for the estimated parameters is essential for obtaining a reasonable estimate and not to stick into an unimportant local minimum. Experiments demonstrate the performance of this optimization technique and show that the choice of the regularization parameter/hyper-parameters is an essential key-point in order to obtain precise motion estimates.

References:

[1] A. Mohammad-Djafari, A full Bayesian approach for inverse problems, Presented at the 15th International Workshop on Maximum Entropy and Bayesian Methods (MaxEnt95), Santa Fe, New Mexico, USA (1995)

[2] Lydia Ng and Victor Solo, A Data-driven Method for Choosing Smoothing Parameters in Optical Flow Problems, Proc. International Conference on Image Processing, Washington, DC, USA (1997), pp. 360-363

Key Words: optical flow, marginalized MAP estimation

UPDATING PROBABILITIES

Ariel Caticha and Adom Giffin
Department of Physics, University at Albany - SUNY, Albany, NY, USA
(e-mail: ariel@albany.edu)

Abstract

The Method of Maximum (relative) Entropy (ME) has been designed for updating from a prior distribution to a posterior distribution when the information being processed is in the form of a constraint on the family of allowed posteriors. This is in contrast with the usual MaxEnt which was designed as a method to assign, and not to update, probabilities. The objective of this paper is to strengthen the ME method in two ways.

In [1] the axioms that define ME have been distilled down to three; here the design is improved by considerably weakening the axiom that refers to independent subsystems. Instead of the old axiom which read: "When a system is composed of subsystems that are *believed* to be independent it should not matter whether the inference procedure treats them separately or jointly" we now modify it by replacing the word 'believed' by the word 'known'. As pointed out by Karbelkar and by Uffink the modified axiom is a much weaker consistency requirement, which, in their view, fails to single out the usual (logarithmic) relative entropy as the unique tool for updating. It merely restricts the form of the entropy to a one-dimensional continuum labeled by a parameter η; the resulting η-entropies are equivalent to the Renyi or the Tsallis entropies. We show that further applications of the *same* modified axiom select a unique, universal value for the parameter η and this value corresponds to the usual (logarithmic) relative entropy. The advantage of our new approach is that it shows precisely how it is that the other η-entropies are ruled out as tools for updating.

Our second concern is mostly pedagogical. It concerns the relation between the ME method and Bayes' rule. We start by drawing the distinction between Bayes' *theorem*, which is a straightforward consequence of the product rule for probabilities, and Bayes' *rule*, which is the actual updating rule. We show that Bayes' rule can be derived as a special case of of the ME method. The virtue of our derivation, which hinges on translating the information in data into constraints that can be processed by ME, is that it is particularly clear. It throws light on Bayes' rule and it shows the complete compatibility of Bayes' updating with ME updating.

References:

[1] A. Caticha, "Relative Entropy and Inductive Inference," in *Bayesian Inference and Maximum Entropy Methods in Science and Engineering*, ed. by G. Erickson and Y. Zhai, AIP Conf. Proc. **707**, 75 (2004) (arXiv.org/abs/physics/0311093).

Key Words: relative entropy, Bayes rule

FROM OBJECTIVE AMPLITUDES TO BAYESIAN PROBABILITIES

Ariel Caticha
Department of Physics, University at Albany - SUNY, Albany, NY, USA
(e-mail: ariel@albany.edu)

Abstract

Many discussions on the foundations of quantum theory start from the abstract mathematical formalism of Hilbert spaces and some ad hoc "postulates" or rules prescribing how the formalism should be used. Their goal is to discover a suitable interpretation.

The Consistent-Amplitude approach to Quantum Theory (CAQT) is different in that it proceeds in the opposite direction: one starts with the interpretation and then derives the mathematical formalism from a set of "reasonable" assumptions. The overall objective is to predict the outcomes of certain idealized experiments on the basis of information about how complicated experimental setups are put together from simpler ones. The theory is, by design, a theory of inference from available information.

The "reasonable" assumptions are four. The first specifies the kind of setups about which we want to make predictions. The second assumption establishes what is the relevant information and how it is codified. It is at this stage that amplitudes and wave functions are introduced as tools for the consistent manipulation of information. The third and fourth assumptions provide the link between the formalism and the actual prediction of experimental outcomes. Although the assumptions do not refer to probabilities, all the elements of quantum theory, including indeterminism and the Born rule, Hilbert spaces, linear and unitary time evolution, are derived.

Within the CAQT approach probabilities are completely Bayesian, and yet, there is nothing subjective about the wave function that conveys the relevant information about the (idealized) experimental setup. The situation is quite analogous to assigning Bayesian probabilities to outcomes of a die toss based on the objective information that the (idealized) die is a perfectly symmetric cube.

Key Words: quantum theory, quantum information theory, Bayesian quantum mechanics

BLIND SOURCE SEPARATION USING MAXIMUM ENTROPY PDF ESTIMATION BASED ON FRACTIONAL MOMENTS

Babak Abbasi Bastami, Hamidreza Amindavar
Amirkabir University of Technology,
Department of Electrical Engineering, Tehran, Iran
(e-mail: babak_bastami@yahoo.com, hamidami@aut.ac.ir)

Abstract

Recovering a set of independent sources which are linearly mixed is the main task of the blind source separation. Utilizing different methods such as infomax principle, mutual information and maximum likelihood leads to simple iterative procedures such as natural gradient algorithms[1]. These algorithms depend on a nonlinear function (known as score or activation function) of source distributions. Since there is no prior knowledge of source distributions, the optimality of the algorithms is based on the choice of a suitable parametric density model.

In this paper, we propose an adaptive optimal score function based on the fractional moments of the sources. In order to obtain a parametric model for the source distributions, we use a few sampled fractional moments to construct the maximum entropy probability density function (PDF) estimation [2]. By applying an optimization method we can obtain the optimal fractional moments that best fit the source distributions. Using the fractional moments instead of the integer moments causes the maximum entropy estimated PDF to converge to the true PDF much faster .

The simulation results show that unlike the most previous proposed models [3] for the nonlinear score function, which are limited to some sorts of source families such as sub-gaussian and super-gaussian or some forms of source distribution models such as generalized gaussian distribution, our new model achieves better results for every source signal without any prior assumption for its randomness behavior.

References:

[1] J. F. Cardoso,“Blind signal separation : Statistical Principles,” *Proc. IEEE*, vol 9, no 10, October 1998.

[2] H. Gzyl, P. N. Inveradi, A. Tagliani, M. Villasana, “Maxentropic solution of fractional moment problems,” *Applied Mathematics and Computation*, 2005.

[3] A. Cichocki, S. Amari, “Adaptive blind signal and image processing ,” *John Wiley and Sons*, 2003.

Key Words: Fractional Moments, Blind Source Separation, Score Function

COMPETITIVE BIDDING IN A CERTAIN CLASS OF AUCTIONS

Mathias Johansson[1,2]
(1) Signals and Systems Group, Uppsala University, Sweden,
e-mail: mj@signal.uu.se
(2) Dirac Research AB, Sweden, e-mail: mj@dirac.se

Abstract

We consider a problem of determining the amount to bid in a certain type of auctions in which customers submit one sealed bid. The bid reflects the price a customer is willing to pay for one unit of the offered goods. The auction is repeated and at each auction each customer requests a certain amount of goods, an amount that we call the capacity of the customer and that varies among customers and over time. At each auction, only the customer with the largest bid-capacity product obtains any goods. The price paid by the winner equals his/her bid-capacity product, and the amount of goods obtained in return equals the winner's capacity. The auction is repeated many times, with only limited information concerning winning bid-capacity products being announced to the customers. This situation is motivated in for example wireless communication networks in which a possible way of obtaining a desired service level is to use dynamic pricing and competitive bidding. In this application, the capacity is typically uncertain when the bid is made. We derive bidding rules and loss functions for a few typical service requirements.

We assume that the auctioneer announces only some limited aggregate statistics from previous auctions. Consequently, we use the maximum entropy principle in assigning probabilities for other customers' bids and capacities.

Our approach is to minimize the expected loss, conditional on the limited information I available to the customer. Let a particular customer u's probability that he or she will have the largest bid-capacity product of all customers be denoted by $P(u \mid I)$. Then $P(u \mid I)$ is equal to the probability that the customer v with the largest bid-capacity product of all other customers has a lower bid-capacity product than customer u. Let q_v denote the bid of v, c_v the corresponding capacity, and $y = q_v c_v$ the largest bid-capacity product among all customers except u. We can then find the probability that u wins as follows: first determine the probability that $y < c_u q_u$ assuming knowledge of c_u, i.e. $\int_0^{c_u q_u} P(y \mid c_u I)dy$. Then multiply this with the probability distribution for c_u given I to obtain the joint probability for c_u and $y < c_u q_u$. Integrating the result over all possible capacities c_u, we have

$$P(u \mid I) = \int P(c_u \mid I) \int_0^{c_u q_u} P(y \mid c_u I) dy dc_u \ . \tag{1}$$

In the full paper, we compute this probability explicitly for some particular states of knowledge I and illustrate how customers behave using the suggested strategy.

PARAMETER ESTIMATION OF ELLIPSOMETRY MEASUREMENTS

Udo v. Toussaint[1], Thomas Schwarz-Selinger[1]
(1) Institute for Plasma Physics, 85748 Garching, Germany
(e-mail: udo.v.toussaint@ipp.mpg.de)

Abstract

Ellipsometry is a unique technique of great sensitivity for in situ non-destructive characterization of surfaces utilizing the change in the state of polarization of a light-wave probe which is extensively used in the semi-conductor industry. To relate ellipsometric measurements to surface properties (as eg layer thickness changes in the range of nm or chemical composition), Bayesian probability theory is used. The parameter estimation process is complicated by the incomplete phase information of the measured data. Examples of 3-D surface reconstructions of samples after plasma exposure demonstrate the tremendous information gain due to the Bayesian analysis.

References:

Key Words: Parameter Estimation, Bayesian Probability Theory

Applied Probability Trust (19 May 2005)

ENTROPY FOR PARETO (IV), BURR, AND ITS ORDER //STATISTICS DISTRIBUTIONS

GHOLAMHOSSEIN YARI,* *Iran University of Science and Technology*

GHOLAMREZA MOHTASHAMI,** *Birjand University*

Abstract

Main result of this paper is to derive the exact analytical expressions of entropy for Pareto, Burr and related distributions. Entropy for kth order statistic corresponding to the random sample size n from these distributions is introduced. These distributions arise as tractable parametric models in reliability, actuarial science, economics, finance and telecommunications. We showed that all the calculations can be obtained from one main dimensional integral whose expression is obtained through some particular change of variables. Indeed, we consider that this calculus technique for that improper integral has its own importance.

Keywords: Gamma and Beta functions; Polygamma functions; Entropy; Order Statistics; Pareto, Burr models.

* Postal address: Iran University of Science and Technology, Narmak, Tehran 16844, Iran.
`email: yari@iust.ac.ir`

** Postal address: Birjand university `email: Gmohtashami@birjand.ac.ir`

CLEARING UP THE MYSTERIES: COMPUTING ON HYPOTHESIS SPACES

Kevin H. Knuth
Department of Physics
University at Albany (SUNY)
Albany NY, 12222, USA
(e-mail: kknuth@albany.edu, http://www.huginn.com/kknuth)

Abstract

We all have become very comfortable with Bayesian probability theory and the interpretation of probabilities as real numbers representing degrees of belief. Indeed this level of comfort was necessary for these methods to become widely accepted. In keeping with Jaynes' original goal of 'Clearing up the Mysteries', I aim to inject some healthy discomfort back into this meeting by closely examining hypothesis spaces and the computations we perform on them. For example, these spaces are not necessarily Boolean spaces. There are two different types of logical *and* operations, one which occurs within a lattice and the other which is induced by the lattice product. Clearly, these details do not upset the Bayesian inferential framework with which we have become so comfortable. Instead, they serve to highlight the fact that even today there remains uncharted territory in the foundation of probability theory.

Key Words: Hypothesis Space, Bayes' Theorem, Boolean Algebra, Lattice Theory, Associativity, Distributivity

AND IF YOU WERE A BAYESIAN WITHOUT KNOWING IT?

Bruno Lecoutre
C.N.R.S. et Université de Rouen, France
(e-mail: bruno.lecoutre@univ-rouen.fr
http://www.univ-rouen.fr/LMRS/Persopage/Lecoutre/Eris)

Abstract

Many statistical users misinterpret the p-values of significance tests as "inverse" probabilities ($1 - p$ is "the probability that the alternative hypothesis is true"). As is the case with significance tests, the frequentist interpretation of a 95% confidence interval involves a long run repetition of the same experiment: in the long run 95% of computed confidence intervals will contain the "true value" of the parameter; each interval in isolation has either a 0 or 100% probability of containing it. Unfortunately treating the data as random even after observation is so strange this "correct" interpretation does not make sense for most users. Ironically it is the interpretation in (Bayesian) terms of "a *fixed* interval having a 95% chance of including the true value of interest" which is the appealing feature of confidence intervals. Moreover, these "heretic" misinterpretations of confidence intervals (and of significance tests) are encouraged by most statistical instructors who tolerate and even use them. For instance Pagano (1990, page 288), in a book which claims the goal of "*understanding statistics*", describes a 95% confidence interval as "an interval such that the probability is 0.95 that the interval contains the population value".

The literature is full of Bayesian interpretations of frequentist p-values and confidence levels. All the attempts to rectify these interpretations have been a loosing battle. In fact such interpretations suggest that most users are likely to be Bayesian "without knowing it" [2] and really want to make a different kind of inference [3].

References:
[1] Pagano, R. R., *Understanding Statistics in the Behavioral Sciences* (1990, 3rd edition), West, St. Paul, MN.
[2] Lecoutre B., Et si vous étiez un bayésien "qui s'ignore"? *La Revue de Modulad* 32 (2005) [http://www-rocq.inria.fr/axis/modulad/archives/numero-32/lecoutre-32/lecoutre-32.pdf].
[3] Lecoutre, B., Lecoutre, M.-P. & Poitevineau, J., Uses, abuses and misuses of significance tests in the scientific community: won't the Bayesian choice be unavoidable? *International Statistical Review* 69 (2001), 399-418.

Key Words: Frequentist probabilities, Bayesian probabilities

PROBABILITY ASSIGNMENT IN A QUANTUM STATISTICAL MODEL

L. F. Lemmens [1], Burhan Bakar[1]
(1) Department of Physics, University of Antwerp, Belgium
(e-mail: lucien.lemmens@ua.ac.be)

Abstract

In ref [1] the evolution of a quantum system, appropriate to describe nano-magnets, is mapped on a Markov process when the system is cooled, the adjoint heating process is obtained using Bayes theorem. Once the mapping is achieved a Markov representation for the evolution with respect to inverse temperature of the quantum system is obtained. The representation can be used to study the probability density of the magnetization. The PDF changes from unimodal to bimodal as a function of the temperature. The change occurs at the so called blocking temperature and depends critically on the initial probability. This probability encodes the multiplicity of the states. The transition from paramagnetic to super-paramagnetic behavior is of importance to enhance the sensitivity of the nano-magnet.

Using the information entropy [2] one can calculate the same PDF without invoking a Markov process. Although the characteristics of the PDF for both calculations are resembling, the numerical values are different: implying that probabilities obtained using the trace and the diagonal elements i.e. the method leading to the information entropy, are not necessarily equal to those derived from the Markov process.

Considering both approaches as a model to assign probabilities, one can use the maximum entropy principle to perform a model selection. A straight forward calculation shows that the entropy obtained in the Markov representation is larger than the information entropy.

References:

[1] Burhan Bakar, and L F Lemmens Phys. Rev. E 71, 046109 (2005) see also cond-mat / 0502277

[2] Alexander Stotland, Andrei A. Pomeransky, Eitan Bachmat, Doron Cohen, Europhysics Letters 67, 700 (2004)

Key Words: Blocking Temperature, information entropy, Markov representation

Maximum likelihood separation of spatially auto-correlated images using a Markov model

Shahram Hosseini[1], Rima Guidara[1], Yannick Deville[1], Christian Jutten[2]
(1) Laboratoire d'Astrophysique de Toulouse-Tarbes (LATT), France
(2) Laboratoire des Images et des Signaux (LIS), Grenoble, France

Abstract

We recently proposed an efficient maximum likelihood approach for blindly separating markovian time series [1]. In the present paper, we extend this idea to bi-dimensional sources (in particular images), where the spatial correlation of each source is described using a 2nd-order Markov model. The idea of using Markov Random Fields (MRF) for image separation has recently been exploited by other authors [2], where the source Probability Density Functions (PDF) are supposed to be known, and are used to choose the Gibbs priors. In the present work, however, we make no assumption about the source PDF so that the method can be applied to any sources. Beginning with the joint PDF of all the observations, and supposing a unilateral 2nd-order Markov model for the sources, we can write down the likelihood function and show that the nondiagonal entries of the separating matrix can be estimated by solving the following estimating equations

$$E[\sum_{k=0}^{1} \sum_{l=-1,k+l\neq-1}^{1} \psi_{s_i}^{(k,l)}(m,n)\hat{s}_j(m-k,n-l)] = 0 \qquad i \neq j$$

where the conditional score functions $\psi_{s_i}^{(k,l)}$ of the estimated sources $\hat{s}_i$ are

$$\psi_{s_i}^{(k,l)}(m,n) = \frac{-\partial \log P_{s_i}(\hat{s}_i(m,n)|\hat{s}_i(m-1,n-1), \hat{s}_i(m-1,n), \hat{s}_i(m-1,n+1), \hat{s}_i(m,n-1))}{\partial s_i(m-k,n-l)}$$

In practice, these functions must be estimated from data in a 5-dimensional space. The nonparametric estimation algorithm used in [1] being very time consuming, we developed a new algorithm using polynomials as score function parametric models. The estimating equations were solved using Newton algorithm. The experiments proved the better performance of our method in comparison to some classical algorithms. The final version of the paper will contain the theoretical details and the experimental results with artificial and real data including astrophysical images.

References:

[1] S. Hosseini, C. Jutten, D.-T. Pham, Markovian source separation, *IEEE Transactions on Signal Processing*, vol. 51, no. 12, pp. 3009-3019, Dec. 2003.

[2] E. E. Kuruoglu, A. Tonazzini and L. Bianchi, Source separation in noisy astrophysical images modelled by markov random fields, *ICIP'04*, pp. 2701-2704.

Key Words: Blind source separation, Markov random fields, Maximum likelihood

Comparing Class Scores in GCSE Modular Science

JASON WELCH, County High School, Leftwich, Cheshire, UK

Abstract

Multiple choice tests are used widely in education and elsewhere. The results of these tests contain information both about the students' knowledge and their ability to guess the answers. This paper describes the use of Bayesian statistical techniques to attempt to 'remove' the guess-work from the results in order to obtain information about the students' underlying knowledge based on our prior knowledge about the structure of the test. The resulting mathematical model allows fair comparisons of the levels of knowledge of groups of students in schools and highlights the flaws in the common practice of analysing these scores using simple averages. It also allows more specific comparisons to be made that are not possible using averages. These comparisons can then inform teaching practice.

Introduction

A multiple choice test provides the student with a number of options from which they are to select the correct answer *e.g.*

The Milky Way is a …

A galaxy

B solar system

C universe

D star

Using such a test to assess knowledge can be problematic not least because the person being tested could guess the correct answer without any understanding of the topic. The literature on multiple-choice testing is wide-ranging but can be broadly categorised into four areas: question writing, administration of tests (electronically), scoring systems and results analysis. The work comes predominantly from higher-education (especially in medicine, law, economics and IT) with contributions from statistics and psychology.

Dirichlet or Potts ?

Ali Mohammad-Djafari
Laboratoire des signaux et systmes (L2S)
Suplec, Plateau de Moulon, 3 rue Joliot Curie
91192 Gif-sur-Yvette, France

Abstract

When modeling the distribution of a set of data $\{x_i, i = 1, \cdots, n\}$ by a mixture of Gaussians (MoG), there are two possibilities: i) the classical one is using a set of parameters which are the proportions α_k, the means μ_k and the variances σ_k^2; ii) the second is to consider the proportions α_k as the probabilities of a hidden variable z whith $\alpha_k = P(z = k)$ and assignining a prior law for z. In the first case a usual prior distribution for α_k is the Dirichlet which account for the fact that $\sum_k \alpha_k = 1$. In the second case, to each data x_i we associate a hidden variable z_i. Then, we have two possibilities: either assuming the variables z_i to be i.i.d. or assigning them a Potts distribution. In this paper we give some details on these models and different algorithms used for their simulation and the estimation of their parameters.

More precisely, in the first case, the assumption is that the data are i.i.d samples from $p(x) = \sum_{k=1} \alpha_k \mathcal{N}(\mu_k, \sigma_k^2)$ and the objective is the estimation of $\theta = \{K, (\alpha_k, \mu_k, \sigma_k^2), k = 1, \cdots, K\}$. In the second case, the assumption is that the data x_i is a sample from $p(x_i|z_i = k) = \mathcal{N}(\mu_k, \sigma_k^2), \forall i$ where the z_i can only take the values $k = 1, \cdots, K$. Then if we assume z_i i.i.d., then the two models become equivalent with $\alpha_k = \frac{1}{n} \sum_{i=1}^{n} \delta(z_i - k)$. But if we assume that there some structure in the hidden variables, we can use the Potts model $p(z_i|z_j, j \neq i) \propto \exp\left\{\gamma \sum_{j \in \mathcal{V}(i)} \delta(z_i - z_j)\right\}$ where $\mathcal{V}(i)$ represents the neighboring elements of i, for example $\mathcal{V}(i) = i - 1$ or $\mathcal{V}(i) = \{i - 1, i + 1\}$ or in cases where i represents the index of a pixel in an image, then $\mathcal{V}(i)$ represents the four nearest neigbors of that pixel. γ is the Potts parameter.

These two models have been used in many data classification or image segmentation where the x_i represents either the grey level or the color components of the pixel i and z_i its class labels. The main objective of an image segmentation algorithm is the estimation of z_i. When the hyperparameters K, $\theta = (\alpha_k, \mu_k, \sigma_k^2), k = 1, \cdots, K$ and *gamma* are not known and have also to be estimated, we say that we are in *totally unsupervised* mode, when are known we are in *totally supervised* mode and we say that we are in *partially supervised* mode when some of those hyperparameters are fixed. In the following, we present some of these methods.

Key Words: Mixture of Gaussians, Dirichlet, Potts, Classification, Segmentation.

ESTIMATION AND DETECTION OF A PERIODIC SIGNAL

D. Aronsson, E. Björnemo, M. johansson
Signals and Systems Group, Uppsala University, Sweden
dar,eb,mj@signal.uu.se

Abstract

Detection and estimation of a periodic signal with an additive disturbance is considered. We study estimation of both the frequency and the shape of the waveform and develop a method based on Fourier series modelling. The method has an advantage over time domain methods such as epoch folding, in that the hypothesis space becomes continuous. Using uninformative priors, the noise variance and the signal shape can be marginalised analytically, and we show that this expression can be evaluated in real time when the data is evenly sampled and does not contain any low frequencies.

We compare our method with other frequency domain methods. Although derived in various different ways, most of these, including our method, have in common that the *cumulative periodogram* plays a central role in the estimation. But there are important differences. Most notable are the different penalty terms on the number of harmonic frequencies. In our case, these enter the equations automatically through the use of probability theory, while in previous methods they need to be introduced in an ad hoc manner. The Bayesian approach in combination with the chosen model structure also allow us to build in prior information about the waveform shape, improving the accuracy of the estimate when such knowledge is available.

SENSOR NETWORK NODE SCHEDULING FOR ESTIMATION OF A CONTINUOUS FIELD

E. Björnemo, M. Johansson, D. Aronsson
Signals and Systems Group, Uppsala University, Sweden
eb,mj,dar@signal.uu.se

Abstract

A wireless sensor network consists of radio-equipped sensors that are spread out in space to perform some network task, such as monitoring or estimating a field quantity. In many sensor networks, the main limitation is the scarce energy resources available at each sensor node. A major issue is therefore optimisation of the activity in the network with respect to energy consumption. We investigate such an energy-limited sensor network, whose purpose is to estimate a continuous field over a certain spatial and temporal range. One way of reducing energy consumption is to utilise knowledge of the field variations to reduce the number of actual measurements and thus not waste energy on measuring quantities that can be inferred with knowledge of related parameters.

We investigate the trade-off between estimation performance and resource cost in terms of energy consumption, and devise a general Bayesian estimation scheme to take advantage of (necessarily incomplete) knowledge of physical properties of the field, such as bounds on time and space variations. Each measurement is taken at a discrete point in space and time and our goal is to infer the entire field over a given time and space horizon. We assume that the position of each node is known and that there is a known node-specific cost associated with each sensor measurement. The central unit schedules sensor measurements according to cost and information gain. We assume simple sensors that only perform the assigned measurements and forward them to the central unit along pre-defined routeS.

We illustrate how different states of uncertainty lead to interesting special cases of the general problem scenario, and discuss relations to Nyquist sampling of a time series of known bandwidth.

Maximum entropy approach to characterization of random media

E. V. Vakarin
UMR 7575 LECA ENSCP-UPMC, 11 rue P. et M. Curie, 75231 Cedex 05, Paris, France
e-mail: vakarin@ccr.jussieu.fr

Characterization of complex disordered media (porous matrices, random networks, etc.) is usually based on an analysis of indirect probes. This is realized through a contact of the medium and a system with a well-defined response function $\theta(\mu|\sigma)$ conditional to the medium state σ. Thus one deals with an inversion of the following integral

$$\theta(\mu) = \int d\sigma f(\sigma)\theta(\mu|\sigma),$$

where $\theta(\mu)$ is an experimental result, $f(\sigma)$ is the desired distribution of some relevant quantity, σ. In many cases the inversion of this integral with respect to $f(\sigma)$ is an "ill-posed" mathematical problem. Therefore, current approaches involve either sophisticated regularization procedures, or a fitting with multiple adjustable parameters. The problem is complicated by the absence of a unique solution and strong sensitivity to the input deviations.

Based on a combination of the statistical thermodynamics and the maximum information principle [1] we propose a complementary approach to this problem. The distributions are calculated through a maximization of the Shannon entropy functional conditioned by the available data $\theta(\mu)$. The scheme is shown to provide an explicit solution and a systematic link between the distribution and the input ($\theta(\mu)$, and $\theta(\mu|\sigma)$). The gained amount of information is shown to be directly related to the probe thermodynamic state $\theta(\mu)$. Several illustrative examples, relevant to adsorption probes are discussed.

[1] E. T. Jaynes, Phys. Rev. **106**, 620 (1957)

INFORMATION-THEORETIC MEASURES OF SOME QUANTUM SYSTEMS

R.J. Yáñez[1,3], J.S. Dehesa[2,3]

(1) Dpto. de Matemática Aplicada,
Universidad de Granada, Spain
(2) Dpto. de Física Moderna,
Universidad de Granada, Spain
(3) Instituto Carlos I de Física Teórica y Computacional,
Universidad de Granada, Spain.
(email: ryanez@ugr.es, dehesa@ugr.es, Fax: 958242862)

Abstract

The distribution of a probability density function all over its domain of definition may be best measured by means of information-theoretic notions of both global (Shannon entropy) and local (Fisher information) characters. These quantities will be here computed for several classical and quantum systems directly from its wave equation. In this communication we shall make emphasis on the following single particle systems: atoms in a spherically symmetric potential, circular membrane and atoms in an external electric field. The extension to multidimensional systems will be also discussed. All these problems require an extensive use of the theory of special functions and orthogonal polynomials.

References:

[1] J. S. Dehesa, A. Martínez-Finkelshtein & V.N. Sorokin, Short-wave asymptotics of the information entropy of a circular membrane, Int. J. Bifurcation and Chaos, 12 (2002) 2387.

[2] J.S. Dehesa, S. López-Rosa, B. Olmos & R.J. Yáñez, The Fisher information of D-dimensional hydrogenic systems in position and momentum spaces, J. Mathematical Physics (2006). Accepted.

Key Words: Shannon entropy, Fisher information, wave equation, circular membrane, quantum mechanics, atoms, atoms in external fields, special functions, orthogonal polynomials

INTEGRATED DATA ANALYSIS: NON-PARAMETRIC PROFILE GRADIENT ESTIMATION

R. Fischer[1], A. Dinklage[2], V. Dose[1]
Max-Planck-Institut für Plasmaphysik, EURATOM Association,
(1) Boltzmannstr. 2, D-85748 Garching, *Germany*
(2) Greifswald Branch, Wendelsteinstr. 1, D-17493 Greifswald, *Germany*
(Rainer.Fischer@ipp.mpg.de)

Abstract

The estimation of distributions and distribution gradients from pointwise measurements of profiles is frequently hampered by measurement errors and lack of information. A combination of measured profile data from heterogeneous experiments is suitable to provide a more reliable data base to decrease the estimation uncertainty by complementary measurements. The Integrated Data Analysis (IDA) concept allows to combine data from different experiments to obtain improved results [1]. Persisting missing information is usually regularized by applying parametric interpolation schemes to fit profiles and derive gradients at the expense of flexibility. The lack of flexibility affects in particular the estimation of profile gradients. The estimation of profile gradient uncertainties is usually not considered. The goal is to reconstruct profiles only from the significant information in the measured data and avoid noise fitting without restricting profiles using parametric functions.

A flexible non-parametric distribution estimation is achieved by using exponential splines. Exponential splines adaptively allow for flexibility in regions where profile data provide detailed information as well as smoothness (cubic splines as limiting case) elsewhere. Regularization parameters as well as number of knots and knot positions are marginalized in the framework of Bayesian probability theory. The resulting posterior probability distribution allows to estimate profiles, profile gradients and their uncertainties in a natural way. An application of exponential splines will be shown on temperature and density profile gradient estimation from an integrated data set measured with different experiments for transport modeling at Wendelstein 7-AS and ASDEX Upgrade.

References:

[1] R. Fischer, A. Dinklage, and E. Pasch, Plasma Phys. Control. Fusion **45** (2003) 1095

[2] V. Dose and A. Menzel, Global Change Biology **10** (2004) 259

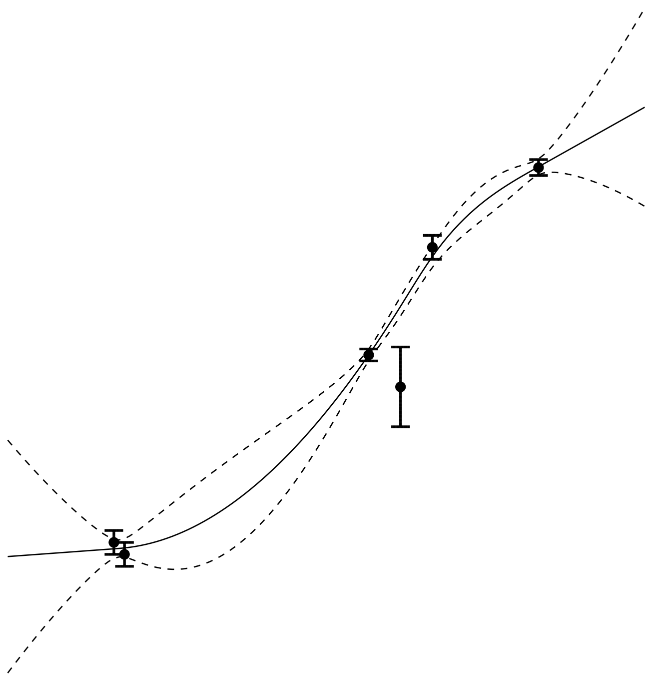
[3] V. Dose and R. Fischer, Bayesian Inference and Maximum Entropy Methods in Science and Engineering, ed. K. Knuth, AIP Conf. Proc. 803 (2005) 67

CALIBRATION and INTERPOLATION

John Skilling
Maximum Entropy Data Consultants Ltd
Kenmare, County Kerry, Ireland
(e-mail: skilling@eircom.net)

Abstract

Interpolation is the problem of fitting a curve $y(x)$ to data (which may be inaccurate), and calibration refers to reading $x(y)$ from this curve. Traditionally, spline interpolation is often used, but this tends to overshoot and ring badly unless the calibration points are evenly spaced. But this restriction should not be needed. Here as elsewhere in science, more data should give better results. This paper presents a free-form probabilistic solution controlled by the degree of curvature of the interpolant. The analysis puts the spline suggestion into context, and generates appropriate uncertainties.



Key Words: Bayesian, calibration, interpolation, free-form.

The Full Bayesian Significance Test for Separate Hypotheses

Julio M. Stern[1], Marcelo S. Lauretto[1],
Silvio R. de Faria Jr.[1], Basilio B. Pereira[2]
(1) University of Sao Paulo, Brazil
(2) Federal University of Rio de Janeiro, Brazil
(e-mail: jstern@ime.usp.br)

Abstract

A typical problem of discriminating between models consists of determining which of m alternative models, $f_k(x, \psi_k)$, more adequately fits or describes a given dataset. In general the parameters ψ_k have distinct dimensions, and the models f_k have distinct (unrelated) functional forms. In this case it is usual to call them "separate" models (or hypotheses). Atkinson [1], although in a different theoretical framework, was the first to analyse this problem using a mixture formulation,
$f(x|w_1 \ldots w_m, \psi_1 \ldots \psi_m) = \sum_{k=1}^{m} w_k f_k(x, \psi_k)$, where $w_k \geq 0$, $\sum_{k=1}^{m} w_k = 1$.

The Full Bayesian Significance Test (FBST) was introduced by Pereira and Stern in 1999 and its invariant formulation was presented by Madruga et al [2]. The FBST was applied in mixture model selection by Lauretto and Stern [3] and performed very well when compared with model-based clustering methods.

In this article we propose the FBST as a robust tool for the test of separate hypotheses, in the context of mixture formulation. Simulated experiments in the Lognormal *versus* Gamma and other classical problems are analysed, where the FBST performance is compared with Bayes Factors [4].

References:

[1] A.C.Atkinson (1970). A Method for Discriminating Between Models. *J. Royal Stat. Soc. B*, 32, 323–354.

[2] M.R.Madruga, C.A.B.Pereira, J.M.Stern (2003). Bayesian Evidence Test for Precise Hypotheses. *Journal of Statistical Planning and Inference*, 117,185–198.

[3] M.S.Lauretto, J.M.Stern (2005). FBST for Mixture Model Selection. Maxent'2005, *AIP Conf. Proc.* 803, 121-128.

[4] B.B.Pereira (2005). Separate Families of Hypotheses. In: Peter Armitage, Theodore Calton (Org.) Encyclopedia of Biostatistics 2 ed. v.7, 4881-4886.

Key Words: Bayes factors, mixture models, separate hypotheses, significance test.

AN ALTERNATIVE APPROACH TO THE PARAMETRIC EMPIRICAL BAYES SELECTION OF WAVELET THRESHOLDS

M. Aminghafari

Dépt. Statistique et Modélisation, Univ. Paris-sud and Dept. of Statistics, Tehran Polytechnic

Mina.Aminghafari@math.u-psud.fr

The prior distribution is the key to the Bayesian inference. Theoretical priors, conjugate priors and estimated priors are different types of prior distributions. Estimating hyper-parameters, i.e., prior distribution parameters, using data is called Parametric Empirical Bayes (PEB), which is used by frequentists more often than by Bayesians [3].

Our purpose in this paper is to provide an alternative approach to the PEB approach to wavelet threshold selection [2,4,5]. We propose an approach to wavelet threshold selection when we have a few prior candidates for the wavelet coefficients. In the other words, instead of PEB estimation of the threshold, we perform a prior selection and then estimate the threshold. A few advantages of the proposed method are given through the examples. We compare the proposed method with the well known methods to the wavelet thresholding [1].

Key Words: Wavelet Thresholding, Parametric Empirical Bayes, Most Powerful Test.

References:

[1] Antoniadis, A., Bigot, J., and Sapatinas, T. (2001) Wavelet estimators in nonparametric regression: a comparative simulation study. *Journal of Statistical Software*, 6:1–83.
[2] Johnstone, I.M. and Silverman, B.W. (2005) Empirical Bayes selection of wavelet thresholds. *The Annals of Statistics*, 33:1700–1752.
[3] Robert, C.P. (2001) *The Bayesian Choice.* 2nd edition, Springer, New York.
[4] Ruggeri, F. and Vidakovic, B. (1999) A Bayesian decision-theoretic approach to the choice of thresholding parameter. *Statistica Sinica*, 9:183–197.
[5] Vidakovic, B. (1999) *Statistical Modeling by Wavelets.* Wiley, New York.

Human Detection using Ultrasonic Doppler Vibrometry

Asif Mehmood[1] , Paul M. Goggans[1], James M. Sabatier[2]
and Alexander E. Ekimov[2]
The University of Mississippi
(1) Electrical Engineering Department
(2) National Center for Physical Acoustics
University, MS 38677
(e-mail: amehmood@olemiss.edu)

Abstract

This paper considers the problem of distinguishing a walking human from other moving or stationary objects in near real-time using Bayesian model comparison and data obtained from an ultrasonic Doppler vibrometer (UDV). Here we describe our initial experimental and analytical work to develop an automated non-invasive model-based approach for recognizing people based on their measured velocity signal while walking. Our experimental set up uses an ultrasonic Doppler vibrometer as a non-contact means for obtaining data related to the velocity of the moving body components. The main advantages of using an ultrasonic vibrational measurement system is high resolution, low cost, and ease of installation. In the UDV an ultrasonic transducer directs a 50 kHz acoustic wave to the moving body surface. The returned acoustic signal, frequency modulated by the velocity of the body components, is received by a co-located transducer whose output is sampled to produce the output data time series. Our experiment is laboratory based and intended to determine basic capabilities. The presence of a characteristic and approximately sinusoidal back motion component in the observed velocity of a walking human is used as a basis for distinguishing a walking human from other moving objects. We have two models, one representing data attributed to human bulk velocity and back velocity and the other model representing a constant (possibly zero) bulk velocity. In the walking human model, the back motion can be linked to whole body motion. For the detection of walking human, we make use of the Bayesian inference approach, where models are compared by computing their posterior odds ratio. In this paper we present parameters estimation results for the walking human model and discuss our initial model selection results.

INTEGRATED BAYESIAN ESTIMATION OF Z_{eff} IN THE TEXTOR TOKAMAK FROM BREMSSTRAHLUNG AND CX IMPURITY DENSITY MEASUREMENTS

G. Verdoolaege[1], M.G. Von Hellermann[2], R. Jaspers[2], W. Biel[3], G. Van Oost[1] and the TEXTOR team
(1) Applied Physics, Ghent University, 9000 Gent, Belgium
(2) FOM-IPP Rijnhuizen, 3430 BE Nieuwegein, The Netherlands
(3) IPP FZ-Jülich GmbH, D-52425 Jülich, Germany

Abstract

The validation of diagnostic data from a nuclear fusion experiment is an important issue. The concept of an Integrated Data Analysis (IDA) allows the consistent estimation of plasma parameters from heterogeneous data sets [1]. Here, the determination of the ion effective charge (Z_{eff}), a critical local measure of impurity concentration, is considered. Several diagnostic methods exist for the determination of Z_{eff}, but the results are in general not in agreement. Moreover, so far none of the available methods has provided a Z_{eff} estimate that is reliable over the entire plasma cross-section, which is at present a real challenge.

In this work, the problem of Z_{eff} estimation is approached from the perspective of IDA, in the framework of Bayesian probability theory. The ultimate goal is the estimation of a full Z_{eff} profile that is consistent both with measured bremsstrahlung emissivities, as well as individual impurity spectral line intensities obtained from Charge Exchange Spectroscopy (CXS). We present an overview of the various uncertainties that enter the calculation of a Z_{eff} profile from bremsstrahlung data on the one hand, and line intensity data on the other hand. These appear at several levels, including the measurement process itself (together with independent electron density and temperature measurements), the inversion procedure (including knowledge of the magnetic equilibrium), the atomic data, the diagnostic calibrations, etc. We discuss a simple Bayesian model permitting the estimation of a central value for Z_{eff} and the electron density n_{e} on TEXTOR from bremsstrahlung emissivity measurements in the visible, and carbon densities derived from CXS. Both the central Z_{eff} and n_{e} are sampled using an MCMC algorithm. Extensions of the model to a full Bayesian analysis, incorporating all critical measurement and model uncertainties, are examined. Relevance to ITER through the pilot active beam experiment on TEXTOR is discussed.

References:

[1] R. Fischer and A. Dinklage, Rev. Sci. Instrum. **75**, 4237 (2004).

PROBING THE COVARIANCE MATRIX

Kenneth M. Hanson
Los Alamos National Laboratory, Los Alamos, NM 87545, USA
(e-mail: kmh@lanl.gov, http://public.lanl.gov/kmh/)

Abstract

Relationships between statistics and physics often provide a deeper understanding that can lead new or improved algorithmic approaches to solving statistics problems. It is well known that the negative logarithm of a probability distribution is analogous to a physical potential. Thus, $\varphi(\boldsymbol{a}) = -\log(p(\boldsymbol{a} \,|\, \boldsymbol{y}))$ is analogous to a potential, where $p(\boldsymbol{a} \,|\, \boldsymbol{y})$ is the posterior, vector $\boldsymbol{a}$ represents the n continuous parameters, and $\boldsymbol{y}$ represents the m measurements. The maximum a posteriori (MAP) solution, $\hat{\boldsymbol{a}}$, which minimizes $\varphi(\boldsymbol{a})$, is frequently chosen as the parameter estimator because it is easier to find than the posterior mean. In many inference problems the posterior can not be stated in analytic form, only evaluated by means of a computational model.

The inference process requires estimates of the uncertainties in , which are related to the width of the posterior, typically characterized in terms of the covariance matrix $\boldsymbol{C}$. Standard approaches to determining $\boldsymbol{C}$ include: 1) sensitivity analysis, 2) Markov chain Monte Carlo, and3) functional analysis. Each of these approaches has its advantages and disadvantages depending on the nature of the problem, for example, the magnitude of n and m and the cost of evaluating the forward model and its sensitivities.

I describe a novel alternative approach that may be advantageous in some situations. In the physics analogy, the notion is to determine the displacement of the equilibrium of the system ($\hat{\boldsymbol{a}}$) under the influence of an external force. The displacement is determined by the curvature (or stiffness) matrix describing the potential around $\hat{\boldsymbol{a}}$. In the inference problem, the idea is to add to $\varphi(\boldsymbol{a})$ a potential that is linear in $\boldsymbol{a}$ and find the new minimizer $\boldsymbol{a}'$. It is easy to show that $\Delta\boldsymbol{a} = \boldsymbol{a}' - \hat{\boldsymbol{a}} = \boldsymbol{C}\boldsymbol{f}$, where $\boldsymbol{f}$ is the force applied to the system; thus, the additional potential is $\boldsymbol{a}^T\boldsymbol{f}$. The force $\boldsymbol{f}$ represents a linear combination of the parameters about which we want to estimate the uncertainty. The variance in the direction of $\boldsymbol{f}$ is proportional to $\boldsymbol{f}^T\boldsymbol{C}\boldsymbol{f} = \boldsymbol{f}^T\Delta\boldsymbol{a}$. Furthermore, the covariance between $\boldsymbol{f}$ and another linear combination of parameters $\boldsymbol{g}$ is $\boldsymbol{g}^T\boldsymbol{C}\boldsymbol{f} = \boldsymbol{g}^T\Delta\boldsymbol{a}$. This approach to uncertainty estimation is most useful in situations in which a) the standard techniques are costly, b) it is relatively easy to find the minimum in $\varphi(\boldsymbol{a})$ and $\varphi(\boldsymbol{a}) + \boldsymbol{a}^T\boldsymbol{f}$, and c) one is interested in the uncertainty wrt. one or a few directions in the parameter space. The useful of this new technique is demonstrated with examples ranging from simple to complicated.

Key Words: covariance estimation, probability potential, posterior stiffness, linear response theory, dissipation-fluctuation relation

A Cyclostationary Bayesian Approach For GPS Signals Delay and Frequency Offset Estimation

M. Sahmoudi and M. G. Amin
Wireless Communications and Positioning Lab
Center for Advanced Communications
Villanova University, PA 19085, USA
(e-mail: mohamed.sahmoudi@villanova.edu)

Abstract

Frequency offset and delay estimation are typically required in the reception of GPS signals. The frequency of the incoming signal can differ from that of the local oscillator frequency due to propagation, Doppler effects and mismatch between the satellites transmitter and the receiver. The delay estimation allows the computation of the distance between the user and the satellite, and with at least four estimates from the different emitters, the positioning can be realized. In GPS applications, conventional positioning techniques are based on the characteristics of the pseudorandom code autocorrelation function. They don't take into account the eventuality presence of interference, multipath reflections and high level non-Gaussian noise. These undesired signals may greatly disturb the measures of the usual Early-Late method [3]. Several techniques has been proposed in the GPS literature to mitigate separately the interference or multipath effect but there is no synchronization method robust to both of them.

Since we have shown in [1, 2] that the cyclostationarity property of the GPS signal can be exploited to improve the synchronization parameters estimation, we propose in this paper a novel method based on the second order cyclic statistics. By considering the sample cyclic autocorrelation function of the GPS signal and the probability distribution of the estimation error, a general linear model formulation of the problem is derived from which the parameters are estimated using the Maximum A Posteriori (MAP) estimator in a Bayesian framework. This approach require only knowledge of the frequency that characterizes the underlying periodicity exhibited by the GPS signal, namely the cycle frequency. Thus we avoid the need for a priori knowledge of interference and noise characteristics (e.g. no gaussianity assumption is needed).

The results demonstrate analytically and also by simulations that greatly improved jammer and interference rejection is achievable by exploiting properly the signal-selectivity properties of the cyclic autocorrelation function of the cyclostationary GPS signals. The proposed method can perform well for multipath signals severely corruptive by noise and interference that exhibit no cyclostationarity with the same cycle frequency.

References:

[1] William A. Gardner (Editor), Cyclostationarity in Communications and Signal Processing, IEEE PESS 1994.

[2] M. Sahmoudi and M. G. Amin, "Unitary Cyclic MUSIC for Direction Finding in GPS Receivers", submitted to the IEEE International Conference SAM 2006.

[3] R. Van Nee, Multipath and multi-transmitter interference in spread-spectrum communication and navigation system, Delft University Press, 1995.

Key Words: GPS, Delay estimation, Interference, Multipath, Cyclostationarity.

A BAYESIAN APPROACH TO INFORMATION RETRIEVAL FROM SETS OF ITEMS

Katherine A. Heller[1], Zoubin Ghahramani[2,3]

(1) Gatsby Computational Neuroscience Unit,
University College London, London WC1N 3AR, UK

(2) Department of Engineering,
University of Cambridge, Cambridge CB2 1PZ, UK
http://learning.eng.cam.ac.uk/zoubin

(3) Machine Learning Department,
Carnegie Mellon University, Pittsburgh, PA, USA

Abstract

We consider the problem of retrieving items from a concept or cluster, given a query consisting of a few items from that cluster. We formulate this as a Bayesian inference problem based on models of human categorization and generalization and describe a very simple algorithm for solving it. Our algorithm ends up with a score which can be evaluated exactly using a single sparse matrix multiplication. This makes it possible to apply the method to retrieval from very large datasets (i.e. millions of items). We evaluate our algorithm on three problems: retrieving movies from a database of movie preferences, finding sets of similar authors based on their word usage in a scientific conference, and finding completions of sets of words appearing in encyclopaedia articles. Compared to "Google Sets", we show that our "Bayesian Sets" retrieval method gives very reasonable set completions. Finally, we show how the Bayesian Sets algorithm can form the basis of a Content-Based Image Retrieval (CBIR) system. I will describe and demonstrate this Bayesian CBIR system and mention a range of other applications of our approach.

Key Words: Information retrieval, Vision, Image Retrieval, Google

THE MINIMUM CROSS-ENTROPY METHOD: A GENERAL ALGORITHM FOR ONE-DIMENSIONAL PROBLEMS

J.C. Cuchí[1], A. Zarzo[2]
(1) Departament d'Enginyeria Agroforestal, ETSEA,
Universitat de Lleida, 25006-Lleida, Spain.
(e-mail: cuchi@eagrof.udl.es)
(2) Depto. Matemática Aplicada, E.T.S. Ingenieros Industriales,
Universidad Politécnica de Madrid, 28006-Madrid, Spain.
and
Instituto Carlos I, Facultad de Ciencias,
Universidad de Granada, 18071-Granada, Spain.
(e-mail: azarzo@etsii.upm.es, http://dmaii.etsii.upm.es/∼azarzo/)

Abstract

We consider in this paper the well known minimum cross-entropy method (MinxEnt) as applied to the problem of constructing approximations to a non-negative function, $f(x)$, when partial information about it is given as a set of constraints of the form:

$$\mu_0 = \int_a^b f(x)dx \quad \mu_j = \int_a^b k_j(x)f(x)dx \quad j = 1, \ldots, n\,,$$

i.e. its normalization, μ_0, and a set of expectation values of certain functions, $k_j(x)$ $(j = 1, \ldots, n)$. On applying the MinxEnt method to this problem, a minimum of the cross-entropy functional

$$\mathcal{E}[f : f_0] = \int_D f(x) \log\left(\frac{f(x)}{f_0(x)}\right) dx$$

has to be computed, where $f_0(x)$ is a prior approximation to $f(x)$, usually obtained from the knowledge of the specific problem in which $f(x)$ and the constraints appears.

One can find in the literature a number of algorithms to deal with this problem which works for some particular situations (see e.g. [1]–[4] among others). Our intention here is to discuss the behavior of the standard optimization methods (Newton, quasi-Newton, ... with line–search of several types) with the aim of developing a general algorithm to solve the minimization problem in the sense that it could be applied to a wide set of densities and constraints, ranging from the discrete to the

BAYESIAN ANALYSIS ON PLASMA CONFINEMENT DATA BASES

R. Preuss, V. Dose, A. Dinklage
Max-Planck-Institut für Plasmaphysik,
EURATOM Association, D-85748 Garching, Germany
(e-mail: preuss@ipp.mpg.de, http://www.ipp.mpg.de/ preuss)

Abstract

Plasma experiments in large fusion devices represent high efforts and costs. Motivated by physical questions many experiments are performed under small variations of settings leading to clustered parameter spaces. Thereby being hardly well conditioned, data bases are fed by mean values and standard deviations of the measured quantities. In the analysis process the data descriptive equation can only approximate the overall confinement behaviour with a power law. Resulting from this approximation and the sparse parameter space outliers and non-Gaussian error statistics are to be expected. The question addressed here is to cope with them in order to develop a robust analysis.

In the sense of the maximum entropy principle a Gaussian results from the assumed knowledge of mean μ and variance σ^2. If we assume instead that the expectation value of $|d_i - \mu|$ is σ_i we obtain a Laplace distribution. It allows for considerable outlier tolerance compared to the Gaussian choice. A combination of the properties of a Gaussian and a Laplace distribution is obtained from an inverse hyperbolic cosine. Its small argument behaviour approximates a Gaussian while the wings of the distribution are Laplace-like.

We examine the impact of the three likelihood functions on Bayesian model comparison which is employed to determine the consistency of confinement data with different physical models. These models derive from different combinations of inclusion/neglect of ion collisions and diamagnetic response of the plasma to be reflected in equations which couple exponents of a power law ansatz. In order to check the validity of the procedure, subsets of confinement data with known behaviour in the above physical properties are tested against a set of models resulting in an acceptance probability of each model under consideration.

Key Words: Laplacian distribution, Hyberbolic likelihood, Model comparison

OVERVIEW OF BAYESIAN INFERENCE, MAXIMUM ENTROPY AND SVM METHODS

Mihai Costache[1], Henri Maitre[1], Mihai Datcu[1 2]
(1) GET, Telecom Paris, 46 rue Barrault, 75013 Paris, France
(2) German Aerospace Center(DLR),DLR-Oberpfaffenhoffen,Germany
(e-mail: costache@tsi.enst.fr, henri.maitre@enst.fr), mihai.datcu@dlr.de

Abstract

Feature selection analyses from the perspective of classification performance conducts to the discussion of the relationships between SVM, Bayesian and Maximum Entropy formalisms. Maximum Entropy discrimination can be seen as a particular case of Bayesian inference, which at its turn can be seen as a regularization approach applicable to SVM. Probability measures can be attached to each feature vector thus, feature selection can be described by a discriminative model over the feature space. Further the probabilistic SVM allows to define a posterior probability model for a classifier. Further, the similarities with the kernels based on Kullback-Leibler divergence can be deduced, thus returning to a MaxEnt similarity.

Key Words: Bayesian Inference, Support Vector Machine, Maximum Entropy

INFORMATION THEORIC FRAMEWORK FOR THE EARTHQUAKE RECURRENCE MODELS : Methodica Firma Per Terra Non-Firma

Özcan Esmer[1]
(1) Middle East Technical University
(e-mail: esmer@metu.edu.tr)

Abstract

This paper has the Great Expectations to end the complaints of the seismologists that the earthquakes are unpredictable and to replace the traditional methodology with the new one based on the Information Theoric Framework. In words of Charles Dickens (1812-1870) It was the best of times,.. .it was the Age of Wisdom, , since we are living in an Extra Ordinary time at the turn of the Millennium, Age of Information...

On 17 August 1999, a destructive magnitude 7.4 earthquake occurred 100 km east of Istanbul on the North Anatolian Fault. What is the probability of an earthquake of M=7.4 will occur before the year 2030 in Istanbul ? A group of seismologists found a 62 15

International Conference during 1-4 Nov. 2005 in Lisbon, on the occasion of the 250th Anniversary of the 1755 Lisbon Earthquake that influenced not only Portugal but the all Europe & North African countries, was to foster an integrated view of global perception of natural disasters. [http://www.lisbon1755.org]

100th Anniversary of the 1906 San Francisco Earthquake to be held during 18-22 April 2006 shall also include the Centennial Meeting of the Seismological Society of America , where the next 10 steps our communities must take to avoid catastrophic disasters. [http://www.1906eqconf.org]

Paper maintains that the attempts to forecast or predict earthquake occurrence can be studied chronologically under 3 types of models :

1-) Models developed between years 1968-1976 can be designated as the First Generation Models that were based on earthquake probabilities independent of time & geographical location. 2-) Second Generation Models during the next two decades introduced the space & time dimension by considering the local geological & seismological conditions in the estimation of random probabilities. 3-) Third Generation Models developed after 2000, in addition to the above considerations, compute probabilities with respect to the interactions between the local stress changes & the occurrence of large & small earthquakes. [1]

In these three types of models above, many researchers have thought that there ought to be some precursory phenomena that could be consistently observed & identified as the basis for making reliable prediction. There was an intense optimism

Bayesiyan Estimation for Updating Coefficients in the Neural Networks

Modjtaba AMIRI-SAZI and Arastoo SHAFAIE

Computer Engineering and IT Department Amirkabir University of Technology
424 Hafez Ave., Tehran, Iran.
{Amiri_moj, Aristo_sh} @yahoo.com

Abstract. Bayesian inference approach is well known to solve inverse problems in signal and image processing. In this paper we focused on source separation that has become one of the most major fields in the signal and image processing. Classical approaches based on Independent component analysis (ICA) and Principal Component Analysis (PCA) suffer from mapping mixed noises to the desired properties of the sources, on the other hand they try to analyze and separation of signals features using the linear approaches that is unsuitable for non-linear nature of the signals. Therefore using of a non-linear analysis of signals component like Neural-Networks is possible and must be considered. We show that a MLP network with Bayesian estimation for updating coefficients can be used in many signal and image processing application and the proposed algorithm can also serve as very good standard for maximize separation regions, Data Mining applications and minimizing Mutual Information.

Keywords: Blind source separation, Bayesian estimation, Independent component analysis (ICA), Principal Component Analysis (PCA), Multi-Layer Perceptron (MLP).

A QUASILINEAR PARABOLIC TYPE VARIATIONAL SOLUTION FOR FOURIER'S IRREVERSIBLE HEAT CONDUCTION PROCESS WITH MINIMUM PRINCIPLES IN CASE OF FINITE SIGNAL RATE

Endre Kiss
Budapest University of Technology and Economics,
Institute of Physics, Department of Chemical Physics
H-1521 Budapest, Budafoki street 8., Hungary
Telephone: (36-1) 463-1341. Fax:(36-1) 463-1896
(e-mail: kiss_ee@goliat.eik.bme.hu)

Abstract

The linear parabolic type PDE for heat conduction process is analyzed. It is an old well-known problem that with constant phenomenological coefficient the signal spreading velocity is infinite for the Fourier heat conduction process. Here is shown a quasilinear solution for this problem with finite signal rate. Connecting to parabolic PDE it is shown the minimum principle solution of Onsager, Prigogine and Gyarmati type for the Fourier irreversible heat conduction process in energy and entropy representation pictures too. For the stationary state of irreversible heat conduction process there is interesting form for the variational minimum solution with the aid of the so-called "naive" variational procedure. This procedure is equivalent with the Euler-Lagrange PDE. So the minimum entropy production or the minimum information loss can be shown with a more general way. As to the phenomenological solutions of quasilinear heat conduction irreversible process the least dissipation of energy in stationary state leads to the different materials in solid state physics, namely to phonon heat conduction(dielectrics and semiconductors with Umklapp processes) and the conductive electrons (in metals) for which heat conduction coefficients variate in hyperbolic way depending on the temperature. Solutions for them are subharmonic type.

ELECTRODE SELECTION FOR NON-INVASIVE FETAL ELECTROCARDIOGRAM EXTRACTION USING MUTUAL INFORMATION CRITERIA

Frédéric Vrins[1], Reza Sameni[2], Fabienne Parmentier[1], Christophe Hérail[3], Vincent Vigneron[3], Michel Verleysen[1], Christian Jutten[2]
(1) UCL, Louvain-La-Neuve, Belgium
(2) LIS, Grenoble, France
(3) LSC, Evry, France
(e-mail: vrins@dice.ucl.ac.be, www.dice.ucl.ac.be/∼vrins)

Abstract

In recent years, *blind source separation* (BSS) techniques have been used as promising approaches for the non-invasive extraction of fetal cardiac signals from maternal abdominal recordings [1]. With the new developments in BSS, it is believed that the complete shape of the fetal ECG should be extractable from a sufficient number of electrodes well-positioned over the abdomen of a pregnant woman. Based on this intuition, in a previous research a multi-channel recording system containing an array of 72 electrodes was developed, which can be placed as a belt of electrodes over the abdomen and the back of a pregnant woman [2].

However, many of the recording channels are contaminated with the maternal ECG noise and contain little information about the fetal ECG. Moreover, the processing of all the different combinations of these electrodes ($72\times71/2$ electrode pairs), can be very time-consuming and inefficient, since a much smaller subset of the electrodes (which can also vary with time depending on the pose of the fetus, shape of the abdomen, or stage of pregnancy), may be sufficient to extract the required 'information'. Based on this idea, in a recent study, an electrode selection strategy was proposed to reject the channels which correspond to the maternal ECG, by minimizing the *mutual information* (MI) between the different electrodes and a reference channel of the maternal ECG [3].

On the other hand according to the *dipole theory* of the cardiac electrical activity, it is known that the electrodes placed on the body surface are recording a projection of the heart's dipole vector depending on their position. This suggests that the different electrodes should be compared with different references (depending on their location), rather than a single reference. In this work the electrode selection algorithm has been improved by using a novel 3-dimensional model of the cardiac dipole vector. Using this model the reference channel has been customized for each of the recording channels and the MI of each channel has been calculated with respect to its own reference. Moreover the channel selection strategy has also been

A FAST METHOD FOR SPARSE COMPONENT ANALYSIS BASED ON ITERATIVE DETECTION-PROJECTION

Arash Ali AMINI [1], Massoud BABAIE-ZADEH [1,2], Christian JUTTEN [2]
(1) Sharif University of Technology, Tehran, Iran
(2) Laboratory of Images and signals, Grenoble, France
(e-mail: mbzadeh@yahoo.com)

Abstract

We introduce a new iterative algorithm for Sparse Component Analysis (SCA). The algorithm is essentially a method to find sparse solutions of underdetermined linear systems of equations. In the SCA context, the method solves the source separation part of the problem, provided that the mixing matrix is known (*i.e.* estimated). The method is not restricted to SCA and may be used in any context in which such a problem arises. For example, it may be used to find sparse decomposition of a signal in an overcomplete dictionary. For the purpose of discussion, however, we will use the SCA notation and terminology. More specifically, we are given the system $\mathbf{x} = \mathbf{A}\mathbf{s}$ where $\mathbf{x}$ is the known $n \times 1$ mixture vector, $\mathbf{s}$ is the unkown $m \times 1$ source vector and $\mathbf{A}$ is the known $n \times m$ mixing matrix. The system is underdetermined, *i.e.* $n < m$. We wish to find the sparsest source vector satisfying the system.

The idea is to first detect which components of the source vector are active, *i.e.* having a considerable value. The test for activity is carried for each component (*i.e.* each source) separately. We will use a Gaussian mixture to model a (sparse) source. It is found that the optimal test for activity of one source requires the knowledge of all the other sources. We will replace those other sources with their estimates obtained from a previous iteration. The suboptimal test for activity of the i-th source then reduces to comparison of an activity function $g_i(\mathbf{x}, \hat{\mathbf{s}})$ against a threshold. After determination of the activity status of all the sources, the new estimate for source vector will be obtained by finding a solution of $\mathbf{x} = \mathbf{A}\mathbf{s}$ which is closest (in 2-norm) to the subspace specified by the detection step. We will call this step "projection into the activity subspace". Explicit solution of the projection step will be given in terms of pseudo-inverses, for the cases of interest. It is found experimentally that repeated use of the two-step iteration, with proper choices of thresholds, quickly yields the sparsest solution for most well-posed problems.

We will compare the performance of the proposed algorithm against the minimum l^1 norm solution obtained by Linear Programming (LP). It is found by experiment that, with the proper choices of thresholds, the algorithm performs nearly two orders of magnitude faster than interior-point LP solvers while providing the same (or better) accuracy. The figure below shows the typical evolution toward

MARS HYPERSPECTRAL DATA PROCESSING USING ICA AND BAYESIAN POSITIVE SOURCE SEPARATION

H. Hauksdottir [1,4], S. Moussaoui[2], F. Schmidt [3], C. Jutten [1,4], D. Brie [2], J. Chanussot [1,4], J.A. Benediktsson [1,4]
(1) Laboratory of Images and Signals (LIS), Grenoble, France
(2) Centre of Research in Automatic (CRAN), Nancy, France
(3) Laboratory of Planetology, Grenoble, France
(4) University of Iceland, Reykjavik, Iceland
(e-mail: hafrun.hauksdottir@lis.inpg.fr, Said.Moussaoui@cran.uhp-nancy.fr)

Abstract

The surface of Mars is currently being mapped with an unprecedented spatial resolution. The high spatial resolution and its spectral range give it the ability to pinpoint chemical species on Mars more accurately than before. The subject of this paper is to present a method to extract this information. The proposed method combines two approaches, Independent Component Analysis (ICA) [1] and Bayesian positive source separation (BPSS) [2]. ICA will be applied iteratively for selection of pixels in independent locations of the images, while spatial reconstruction SNR will be used to check whether all the regions are accounted for. BPSS is then applied for the estimation of the pure constituent spectra and their abundances. The hyperspectral images are collected with the OMEGA instrument (Observatoire pour la Minralogie, l'Eau, les Glaces et l'Activit), which is a spectrometer boarded on the European Space Agency Mars Express mission and collects 256 images in the infrared spectral region from 0.926 to 5.108 μm with a resolution of 0.014 μm roughly.

As solar light incident to a planetary surface is partially reflected back by interaction with the different constituents, the analysis of reflectance spectra may allow the identification and the quantification of the chemical species present at the surface of Mars. For the linear model the measured spectra is assumed to be a linear mixture of the reflectance spectra, which is the case for geographical mixture of chemical species on the surface. The actual sources are correlated and thus the fundemental assumption of independence in ICA is not satisfied and therefore ICA is not an adequate method for the unmixing. Moreover, it is important to take into account the positivity constraint of both sources and mixing coefficients. This draws attention to Bayesian approach which is able to manage priors such as positivity, but there we face the problem of high computation time when dealing with vast amount of data (more than 30.000 pixels). In this paper, we propose to combine spatial ICA and

Bayesian approximation for physical inverse problem

Roy

Abstract

Here we focus on simulation-based Bayesian inference from electrical impedance tomography (EIT) data. We image an unknown convex polygonal insulating inclusion within an object, made of otherwise conducting material, using current/ voltage measurements on the surface of the object. This kind of problem can be classified as an inverse problem for non-invasive imaging. In the forward map we solve a partial differential equation (PDE) subject to boundary conditions. The statistical inverse problem is to summarize the posterior distribution of conductance at all points within the imaged object given the current applied on the boundary of the object and the corresponding boundary potentials. We apply an MCMC approach to the EIT inverse problem and demonstrate it on noisy simulated data. We will also talk about the coupling separation scheme to analyse the computational approximation of the likelihood.

EMPIRICAL MAXIMUM ENTROPY METHOD

M. Grendar[1] and G. Judge[2]
(1) Bel University and Slovak Academy of Sciences, Slovakia
(2) University of California, Berkeley, USA
(e-mail: umergren@savba.sk)

Abstract

A method, which we suggest to call the Empirical Maximum Entropy method, is implicitly present at Maximum Entropy Empirical Likelihood method [1], as its special, non-parametric case. From this vantage point we will survey the empirical approach to estimation; cf. [1], [2], [3], [4], [5].

References:

[1] R. Mittelhammer, G. Judge and D. Miller, Econometric Foundations, CUP, 2000.

[2] A. Owen, Empirical Likelihood, Chapman-Hall/CRC, 2001.

[3] J. Qin and J. Lawless, Empirical Likelihood and General Estimating Equations, Ann. Statist., 22, 300-325, 1994.

[4] Y. Kitamura and M. Stutzer, An information-theoretic alternative to Generalized Method of Moments estimation, Econometrica, 65, 861-874, 1997.

[5] M. Grendar and G. Judge, Large Deviations theory and Empirical Estimator choice, ARE Berkeley, 2006. http://repositories.cdlib.org/are_ucb/1012

Key Words: Empirical Estimation, Estimating Equations, Criterion Choice Problem

Wavelet-Based SAR Images Despeckling Using Estimation Theory

H. Abrishami Moghaddam, M.J. Valadan Zoj, M. Dehghani,
K.N. Toosi University of Technology, Tehran, Iran

Abstract

In this paper, an improved speckle noise reduction method is presented based on wavelet transform. A 2D circularly symmetric Gaussian function is found to be the best model fitted to the speckle noise pattern cross-section in the logarithmically transformed noisy image [1]. Therefore, a Gaussian low pass filter using a trous algorithm has been used to decompose the logarithmically transformed image. The wavelet coefficients of the signal and noise are modeled using alpha-stable and Gaussian distribution functions, respectively. A Bayesian estimator is then applied to the wavelet coefficients based on these distribution functions as a priori information to estimate the best value for the noise-free signal. Quantitative and qualitative comparisons of the results obtained by the new method with the results achieved from the other speckle noise reduction techniques [2] demonstrated its higher performance for speckle reduction in SAR images.

References:

[1] H. Abrishami, M.J. Valadan Zoj, M. Dehghani, "*Bayesian-based Despeckling in Wavelet Domain Using a trous algorithm*", Proceedings of the 20th International Congress on Photogrammetry and Remote Sensing ISPRS2004, Istanbul, Turkey, July 2004.

[2] Achim, A., A. Bezerianos, and P. Tsakalides, 2001. Novel Bayesian Multiscale for speckle Removal in Medical Ultrasound Images, *IEEE Trans. Med. Imaging Journal,* 20(8):772-783.

Keywords: Speckle Noise, Synthetic Aperture Radar, Bayesian Estimator, Wavelet Transform

RIEMANNIAN OPTIMIZATION METHOD ON THE GENERALIZED FLAG MANIFOLD FOR COMPLEX AND SUBSPACE ICA

Yasunori Nishimori[1], Shotaro Akaho[1], and Mark D. Plumbley[2]
(1) National Institute of Advanced Industrial Science and Technology
AIST Central 2, 1-1-1, Umezono, Tsukuba, Ibaraki 305-8568, Japan
(2) Dep. of Electronic Engineering, Queen Mary University of London
Mile End Road, London E1 4NS, UK
(e-mail: y.nishimori@aist.go.jp)

Abstract

Independent component analysis (ICA) can be solved in two steps: whitening followed by orthogonal rotation. Thus ICA can be tackled by an optimization on the manifold of orthogonal matrices. Recently researchers have investigated the use of manifolds and Lie group methods for ICA and other signal processing tasks, including the Stiefel and the Grassmann manifolds. The aim of this paper is to introduce a new class of manifold: the *generalized flag manifold*. The generalized flag manifold is a set of orthogonal subspaces and includes the Stiefel and the Grassmann manifolds as special cases. This new manifold naturally arises when we relax the condition of ICA and consider subspace ICA. Subspace ICA assumes the source signal s is decomposed into d-tuples where signals within a particular tuple are allowed to be dependent on each other, while signals belonging to different tuples are statistically independent. Then the manifold of candidate matrices is no longer just the Stiefel manifold. The statistical dependence of signals within tuples gives the manifold an additional symmetry, which makes it into the generalized flag manifold. Moreover, the demixing matrix for complex ICA is a unitary matrix as ordinary ICA, and the pair of the real and imaginary parts of each column vector of the unitary matrix forms a 2-dimensional real subspace which is orthogonal to each other. Therefore complex ICA can also be tackled by an optimization on this flag manifold. We extend the Riemannian optimization method to the flag manifold utilizing our previous geodesic formula for the Stiefel manifold [1], and based on it propose a new learning algorithm for complex and subspace ICA. Simulations validate the effectiveness of our method. A part of this work was first presented in [2].

References:

[1] Y. Nishimori and S. Akaho, Learning Algorithms Utilizing Quasi-Geodesic Flows on the Stiefel Manifold, *Neurocomputing*, **67**, pp.106-135, 2005.

[2] Y. Nishimori, S. Akaho, and M D. Plumbley Riemannian Optimization Method on the Flag Manifold for Independent Subspace Analysis, *Proceedings of 6th International Conference ICA2006*, pp.295-302, 2006.

ENTROPIC INFERENCE FOR ASSIGNING PROBABILITIES: SOME DIFFICULTIES IN AXIOMATICS AND APPLICATIONS

Alberto Solana-Ortega, Vicente Solana
Instituto de Matematicas y Fisica Fundamental, CSIC, Spain
(e-mail: albsolana@imaff.cfmac.csic.es)

Abstract

The question how to assign probabilities is inescapable in order to develop a theory of plausible inference, yet it is an extraordinarily difficult one. Among the most attractive candidate procedures for this task we find entropic methods, characterized by the extremization of an entropy functional subject to probability constraints representing available information. Notwithstanding several precedents in concrete disciplines, the first proposal of an entropy method as a general scientific inference procedure was due to Jaynes, in the form of the Principle of Maximum Entropy, with extensions in the method of minimum relative entropy of Kullback, and other generalized formulations.

Here we briefly review the different interpretations and uses of entropy methods. Likewise, we examine the various justifications that have been put forward to support them, in particular the appealing attempts to axiomatically derive a unique mathematical expression for entropy procedures in compliance with consistency requirements. In addition to summarizing the state of the art of the rational foundations for entropic inference, which remain controversial and open, we underline the major difficulties analysts encounter in practice when trying to apply these methods. They are: the selection of an entropy functional, the choice of constraints and the selection of reference measures. The latter issues, about which entropy formalisms remain silent, constitute more than just practical obstacles, and ultimately manifest the incompleteness of inference theory.

In our opinion progress can only be attained with a change of perspective. We propose a logical viewpoint of plausible inference, understood not merely as an optimization problem starting from given probability inputs, but as the representation or encoding of the knowledge that basic evidences and other higher-order information or assumptions provide about conjectures. More specifically, a novel generic scheme for inference is presented, which considers two stages and three inference levels.

MAXENT PRINCIPLE FOR HANDLING UNCERTAINTY WITH QUALITATIVE VALUES

Michele Pappalardo

Department of Mechanical Engineering ,University of Salerno
Salerno, Italy
mpappalrdo@unisa.it

Abstract

A method for handling data in the presence of uncertainty with qualitative values is the theory of Dempster-Shafer. The DS theory is a method for reasoning under uncertainty. The idea of upper and lower theory, include the Bayesian probability as special case, and introduce the *belief function* as lower probabilities and the *plausibility function* as upper probabilities. Here we are interested in applying this theory when the numerical information required by Bayesian methods are not available. The numerical measures in presence of uncertainty may be assigned to a set of propositions as well as to a single proposition. The probabilities are apportioned to subsets and the mass v_i can move over each element.

Let the finite non empty set $=\{x_1,..x_n\}$ be the *frame of discernment* which is the set of all the hypothesis. The basic probability is assigned in the range $[0,1]$ to the 2^n subset of consisting of a singleton or conjunction of singleton of *n* elements x_i . The lower probability $P_{¿}(A_j)$ is defined as $P_{¿}(A_j)=\sum_{A_{j\subseteq ¿A_i}} m(A_j)$.

¿

And the upper probability $P^{¿}(A_j)$ is defined as $P^{¿}(A_j)=1-\sum_{A_{j\subseteq ¿A_i}} m(A_j)$ .The $m(A_i)$ values are the

¿

independent basic values of probability inferred on each subset A_i . The *evidential interval* that provides a lower and upper bound is $EI=[Bl(M),Pl(M)]$. If m_1 and m_2 are the independent basic probabilities from the independent evidence, and $\{A_{1i}\}$ and $\{A_{2j}\}$ the sets of focal points, then the theorem of Shafer gives the rule of combination. Let m_1 and m_2 two independent basic probabilities from the independent evidence.

1

Particle Filtering on Riemannian Manifolds

Hichem Snoussi and Ali Mohammad-Djafari

Particle filtering is an approximate Monte Carlo method implementing the Bayesian Sequential Estimation. It consists in online estimating the a posteriori distribution of the system state given a flow of observed data. The popularity of the particle filter method stems from its simplicity and flexibility to deal with non linear/non Gaussian dynamical models. However, this method suffers from the curse of dimensionality. In general, the system state lies in a constrained subspace which dimension is much lower than the whole space dimension. In this contribution, we propose an implementation of the particle filter with the constraint that the system state lies in a low dimensional Riemannian manifold. The sequential Bayesian updating consists in drawing state samples while moving on the manifold geodesics. We illustrate the effectiveness of the proposed solution on synthetic examples and we show that it compares favorably with classical unconstrained particle filter.

EXTRINSIC GEOMETRICAL METHODS FOR NEURAL BLIND DECONVOLUTION

Simone Fiori
Dipartimento di Elettronica, Intelligenza Artificiale e Telecomunicazioni,
Università Politecnica delle Marche, Italy
(e-mail: fiori@deit.univpm.it, http://www.deit.univpm.it/fiori)

Abstract

The present contribution proposes a Riemannian-gradient-based and a projection-based learning algorithms over a curved parameter space for single-neuron learning. We consider the 'blind deconvolution' signal processing problem using a single neuron model. The learning rule naturally arises as a via criterion-function minimization over the unitary hyper-sphere. We consider the blind deconvolution performances of the two algorithms as well as their computational burden and numerical features.

Why We Should Think of Quantum Probabilities as Bayesian Probabilities

Carlton M. Caves
Department of Physics and Astronomy
1 University of New Mexico
Albuquerque, New Mexico 87131-0001, USA
(e-mail: caves@info.phys.unm.edu
http://info.phys.unm.edu/∼caves)

Abstract

In a realistic, deterministic world, it is easy to argue that all probabilities are subjective Bayesian, i.e., measures of degree of belief. It's so easy, in fact, that the primary refuge for objective interpretations of probability lies in the radically nondeterministic world of quantum mechanics. The objectivist asks, "How can probabilities that are prescribed by physical law be anything but objective?" In this talk I will argue nonetheless that quantum probabilities, even those associated with pure quantum states, are best thought of as being subjective Bayesian probabilities. The viewpoint that emerges from pursuing this line of argument is called the *Bayesian interpretation of quantum mechanics*.

I will review the major arguments for viewing quantum probabilities as subjective, coming from the indistinguishability of quantum states, the apparent nonlocality of entanglement, and the nonuniqueness of ensemble decompositions of mixed states. I will discuss how the quantum de Finetti representation theorem provides a tool for banishing the notion of an unknown quantum state—and a practical tool in quantum cryptography. Finally, I will present an argument, based on the notion of inside information, which clarifies why it seems that the outcomes of measurements on a system in a pure quantum state are more random than a classical random process.

I will conclude with a summary of the Bayesian interpretation of quantum mechanics, including my view of what is objective in a quantum description of the physical world.

Key Words: Quantum probabilities, Bayesian probabilities, Bayesian interpretation of quantum mechanics

INTRODUCTION TO QUANTUM COMPUTATION

Carlton M. Caves
Department of Physics and Astronomy
1 University of New Mexico
Albuquerque, New Mexico 87131-0001, USA
(e-mail: caves@info.phys.unm.edu
http://info.phys.unm.edu/∼caves)

Abstract

Quantum computers are believed to perform some computations, such as factoring large composite numbers, exponentially faster than classical computers. In this talk I will introduce the concepts of quantum computation. Topics to be discussed include how qubits supplant bits, quantum gates and circuits, simple quantum algorithms, quantum error correction, and the requirements for physical implementation of a quantum computer.

Key Words: Quantum computation, qubit, quantum circuit, quantum error correction

THE PARAMETER-BASED FISHER INFORMATION OF RAKHMANOV-BORN DENSITY OF ORTHOGONAL POLYNOMIALS AND QUANTUM SYSTEMS

B. Olmos[1,2], J.S. Dehesa[1,2], R.J. Yáñez[1,3]
(1) Instituto Carlos I de Física Teórica y Computacional,
Universidad de Granada, 18071-Granada, Spain
(2) Departamento de Física Moderna,
Universidad de Granada, 18071-Granada, Spain
(3) Departamento de Matemática Aplicada,
Universidad de Granada, 18071-Granada, Spain
(e-mail: beaos@correo.ugr.es)

Abstract

The ground and excited states of physical systems are described by means of the Born quantum-mechanical probability density $\rho(x \mid \theta)$, which for single-particle systems is equal to the squared wavefunction of the states. Often the physical wavefunctions are given by means of known special functions of the mathematical physics and applied mathematics, and particularly the orthogonal hypergeometric polynomials depending on the parameter θ. Then, the physical Born probability density reduces to the so-called Rakhmanov probability density of the orthogonal polynomials. Here, we calculate the explicit expression for the Fisher information with respect to the parameter θ (not necessarily of locality character) for all the classical orthogonal polynomials in a closed and explicit form. Applications to various specific quantum systems will be described in detail.

References:

[1] J.S. Dehesa, B. Olmos and R.J. Yáñez, *Parameter-based Fisher's information of orthogonal polynomials*, J. Comp. Appl. Math., Preprint (2006)

[2] J.S. Dehesa, S. López-Rosa, B. Olmos and R.J. Yáñez, *The Fisher information of D-dimensional hydrogenic systems in position and momentum spaces*, J. Mathematical Physics, (2006). Accepted

Spreading properties of extreme entropy distributions

S. López-Rosa[1,2], J. C. Angulo[1,2], J. S. Dehesa[1,2]
(1) Instituto Carlos I de Física Teórica y Computacional,
Universidad de Granada, 18071-Granada, Spain
(2) Departamento Física Moderna,
Universidad de Granada, 18071-Granada, Spain
(e-mail: slopez@ugr.es)

Abstract

The extremization of the information-theoretic measures which describe the spreading of the physical states of natural systems gives rise to their fundamental wave equation and/or conservation laws. This is the case not only for the Shannon entropy [1] but also for the Fisher information [2] and the Tsallis entropy [3]. The associated extremun entropy distribution are often known for some given constraints. Here, we carry out a relative comparation of the spreading properties of these distributions for a given similar set of constraints. Some specific applications will be discussed in detail.

References:

[1] See e.g. N. Wu, The Maximum Entropy Method, Edit. Springer-Verlag (1997)

[2] B. R. Frieden, Physics from Fisher Information, Edit. Cambridge University Press (1998)

[3] J. P. Boon and C. Tsallis (edit.), Nonextensive Statistical Mechanics, Europhysics News 36 (2005)

COMPARING CLASSIC, BAYES AND PARAMETRIC EMPIRICAL BAYES RISKS

A. Mohammadpour[1,2], and A. Mohammad-Djafari[2]
(1) Amirkabir University of Technology, Tehran Polytechnic, Iran
(2) L2S, CNRS Supélec Univ. Paris-sud, Gif-sur-Yvette, France

The assumptions on the structure of data or modelling data is a basic problem in the statistical decision making. In this paper we compare the estimator's risks in the different frameworks, with respect to false assumption(s) on the structure of data. More precisely, we compare the robustness of classic, Bayes and parametric empirical Bayes estimators with respect to the prior choice.

Consider independent sample $(X_1, \theta_1), \ldots, (X_n, \theta_n)$ of (X, θ), where $X_i|\theta_i$ has probability density function (pdf) $f(x_i|\theta_i)$ and consider the following cases:

1. In the classical statistics, the prior $\pi(\theta_i)$ is a dirac function, centered on the unknown value θ. That is, $X_1, \ldots, X_n$ are independent and identically distributed random variables with pdf $f(x_i|\theta)$.

2. In the Bayesian case, the prior pdf, $\pi(\theta)$ is no more a degenerate dirac function but is assumed to be perfectly known. In this case, we have only a random parameter θ, i.e. $\theta = \theta_1 = \ldots = \theta_n$, and so the sample $(X_1, \theta), \ldots, (X_n, \theta)$ is not independent and as a result the marginal distributions of X_is are not independent. However, conditional on θ, X_is are independent and have a common pdf $f(x_i|\theta)$.

3. In the parametric empirical Bayes framework, θ_is have pdf $\pi(\theta_i|\tau)$ with the hyper-parameter τ. Conditional on θ_i, X_i has a pdf $f(x_i|\theta_i)$. The common marginal pdf of each X_i is
$$m(x_i|\tau) = \int_{-\infty}^{\infty} f(x_i|\theta_i)\ \pi(\theta_i|\tau)\ \mathrm{d}\theta_i,$$
and $X_1, \ldots, X_n$ are independent.

In this paper, we compare the robustness of estimators in the above frameworks with respect to the false model on data by a Monte Carlo study.

Key Words: Calssic, Bayes and empirical Bayes estimators.

References:

J. O. Berger. *Statistical Decision Theory.* Springer, New York, 1980.
J. M. Bernardo and A. F. M. Smith, *Bayesian Theory.* Wiley, Chichester, UK, 1994.
C. P. Robert, *The Bayesian Choice.* 2nd edition, Springer, New York, 2001.

Maximum Entropy and Bayesian inference: Where do we stand and where do we go?

Ali Mohammad-Djafari
Laboratoire des signaux et systmes(L2S),
UMR 8506 du CNRS-Supélec-Univ. Paris-sud,
Plateau de Moulon, 91192 Gif-sur-Yvette, France

In this tutorial talk, I will first review the main notions of Uncertainty, Random variable, Probabilty distribution, Information and Entropy. Then, we will consider the following main questions in any inference method:
1) Assigning a (prior) probability law to a quantity to represent our knowledge about it,
2) Updating the probability laws when there is new piece of information, and
3) Extracting quantitative estimates from a (posterior) probabilty law

For the first, I will mainly present the Maximum Entropy Principle (MEP).

For the second, we have two tools:
1) Maximising the relative entropy or equivalently minimizing the Kullbak-Leibler discrepency measure, and
2) The Bayes rule.
We will precise the appropriate situations to use them as well as their possible links.

For the third problem, we will see that, even if it can be handeled through the decision theory, the choice of an utility function may depend on the two previous tools used to arrive at that posterior probability.

Finally, these points will be more illustrated through examples of inference methods for some inverse problems such as image restoration or blind sources separation.

Key Words: Uncertainty, Probabilty distribution, Information and Entropy, Maximum Entropy Principle, Bayesian inference, Decision theory

Information theory based inference in the Bayesian context: applications for semantic image coding

Mihai Datcu
German Aerospace Centre (DLR) and GET, Telecom Paris
D-82234 Oberfaffenhaffen

Abstract

Traditionally Information Theory focused to applications in communications, it refers mainly to coding, transmission, or compression of signals. However, implicitly, from its very beginning, information theory closely related to statistics and machine learning. Thus, many other fields like stochastic inference, estimation and decision theory, optimisation, communication or knowledge representation benefit from basic results from information theory. The goal of the tutorial is to overview new applications and new developments in information theory relevant to inference, as well as general methods for information processing and understanding.

The topics envisaged are: applications and extensions of Rate-Distortion theory, the methods of Information Bottleneck, The links to Bayesian, MDL and related methods, information and/or complexity based estimation, and inference.

The lecture will focus on specific methods for: image understanding, image semantic coding, image indexing and information mining. Proposing methods to distinguish, signs and symbols and understand significance.

The possible applications are search engines in large satellite image archives, picture archiving and communication systems (PACS) for use in medical science, or multimedia systems.

1

GraphMaxEnt

Arthur Ramer

School of Computer Science and Engineering
University of New South Wales
Sydney 2032, Australia

Abstract. Assume a undirected graph G on a finite domain X and a probability distribution P on X. Graph entropy, defined in terms of the vertex packing polytope, is recast as

$$H(G,P) = \min_R - \sum p_i \log \mathrm{Pl}^{(R)}(x_i),$$

where plausibility $Pl^{(R)}$ is defined wrt probability distribution R on the stable sets Y (independent subsets of G)

$$\mathrm{Pl}^{(R)}(x) = \sum_{Y:x\in Y} R(Y).$$

The plausibility which obtains from the minimising R is called the *plausibility wrt* P on X

$$\mathrm{Pl}_P(x) = \mathrm{Pl}^{(R)}(x), \quad R = \arg\min H(G,P).$$

It serves to define the *graph information distance*

$$D(G,Q\|P) = \sum q_i \log \frac{\mathrm{Pl}_Q(x_i)}{\mathrm{Pl}_P(x_i)}$$

for two distributions Q and P on X, given a (fixed) graph structure G. It is straightforward to offer a similar definition wrt the change of G, though useful results require some restrictions on that change.

One verifies the usual properties of additivity and subadditivity wrt weak products of the supporting graphs. The method of GraphMaxEnt can be formulated accordingly. It is postulated it admits an axiomatisation akin to that for MaxEnt.

Applied to probability kinematics, it permits resolving several problems arising from the AGM belief revision. One obtains

- imaging (‘nonproportional conditioning’) as minimisation of graph information distance
- Jeffrey-like imaging
- inverse imaging

Applied to trust updating, based on reported experiences, one obtains

- recognition of repeated reports
- recognition of dependent reports

Several further directions are being pursued. On the computational side

- defining DigraphEnt - entropy over directed graphs, or, at least, DAG’s
- constructing continuous domain analogs to $H(G,P)$ and $D(G,Q\|P)$
- use of entropy generating functions

On the foundational side, the main question is physical interpretation of graph entropy. It is hoped that a ‘virtual reality’ can be produced, one that assigns a statistical mechanical meaning to probability kinematics on graph structures.

Key Words: Graph entropy, probability kinematics, AGM model, belief revision, inverse conditioning.