Subjective Bayes

2009

EPSRC
Engineering and Physical Sciences Research Council

THE UNIVERSITY OF WARWICK

HIGHER EDUCATION hefce FUNDING COUNCIL FOR ENGLAND

SUBJECTIVE BAYES

WARWICK 14-16 DECEMBER, 2009

Organisers: Jim Smith, Tony O'Hagan and John Aston

Programme

Monday 14 December

12.00 - 14.00 *Registration and Buffet Lunch - Concourse of the Maths Building*

14.00 - 14.10 Welcome and Arrangements

14.10 - 15.10 Roger Cooke

15.10 - 15.40 Jesus Rios

15.45 - 16.15 *Refreshments - Concourse of the Maths Building*

16.15 - 17.15 Glenn Shafer

17.30 - 19.00 *Poster session (wine and nibbles) - Concourse of the Maths Building*

Roman Belavkin
Adam Binch
Pawel Chwalinski
Alireza Daneshkhah
Radboud Duintjer Tebbens
Guy Freeman
Silvia Liverani
Artemis Parvizi
Richard Wilkinson
Stan Yip

19.00 - 20.30 *Dinner - Rootes Hall*

Tuesday 15 December

9.00 - 10.00 Robert Winkler

10.00 - 10.30 Paul Garthwaite

10.30 - 11.00 *Refreshments - Concourse of the Maths Building*

11.00 – 12:00 Teddy Seidenfeld

12.00 - 12.30 Fernando Vieira Bonassi

12.30 - 14.00 *Buffet lunch - Concourse of the Maths Building*

14.00 – 15.00 Jonathan Rougier

15.00 - 15.30 Leanna House

15.30 - 16.00 *Refreshments - Concourse of the Maths Building*

16.00 - 16.30 Michail Papathomas

16.30 - 17.30 Jim Berger

19.00 *Workshop Dinner - Main Campus*

Wednesday 16 December

9.30 - 10.30 Nick Chater

10.30 – 11.00 Simon French

11.00 - 11.30 *Refreshments - Concourse of the Maths Building*

11.30 - 12.30 Lawrence Phillips

12.30 - 13.00 Conclusions and Discussions

13.00 - 14.00 *Buffet lunch - Concourse of the Maths Building*

Abstracts (Alphabetical by Author)

Roman Belavkin
Middlesex University

Geometric Analysis and Learning

Information is used to define topology on probability measures, which are represented in a vector space. Points in this quasi-metric space are used to represent states of subjective knowledge of a learning system, and an evolution of such a system corresponds to an information trajectory. This geometric approach facilitates the analysis of learning and adaptive systems. In particular, solutions to the information value problem define the optimal trajectory with respect to a utility function. The optimal trajectory is continuous in the defined information topology, and path integrals along the optimal trajectory define the optimal expected utility and information bounds. Parametrisation of the optimal trajectory defines the optimal `temperature' of the exploration-exploitation balance. The theory is demonstrated in a computational experiment.

Jim Berger
Duke University

Bayesian Adjustment for Multiplicity

Issues of multiplicity in testing are increasingly being encountered in a wide range of disciplines, as the growing complexity of data allows for consideration of a multitude of possible hypotheses (e.g., does gene xyz affect condition abc); failure to properly adjust for multiplicities is possibly to blame for the apparently increasing lack of reproducibility in science. Bayesian adjustment for multiplicity is enormously powerful, in that it occurs through the prior probabilities assigned to models/hypotheses. It is, hence, independent of the error structure of the data, the main obstacle to adjustment for multiplicity in classical statistics.

Not all assignments of prior probabilities adjust for multiplicity, and assignments in huge model spaces typically require a mix of subjective assignment and appropriate hierarchical modeling. These issues will be reviewed through a variety of examples. If time permits, some surprising issues will also be discussed, such as the fact that empirical Bayesian approaches to multiplicity adjustment can be seriously flawed.

Adam Binch
University of Sheffield

Bayesian Models of Visual Perception

The two dimensional retinal image constrains, but does not specify, the structure of a scene, and therefore represents an ill-posed problem. The Bayesian framework provides an explanation of how perceptual systems find acceptable solutions to such problems by making use of prior assumptions about the nature of the physical world. That is, sensory information (the probability of which is determined by a likelihood function) is combined with information from previous experience (a prior distribution) to yield a perception, which is determined by the posterior distribution (the product of the likelihood and prior).

The goal of this research is to show that these principles can provide a robust quantitative account of the general methods by which the brain performs computations. This research programme is based on the assumption that the brain's processing of visual information provides a key for understanding the general principles of neural computation.

Fernando Bonassi
Duke University

Bayesian Statistics, De Finetti and The Gambler's Fallacy

We study the problem of prediction in sequences of binary random variables. Models are studied for this kind of situation and then considered vis-a-vis the Gambler's Fallacy - a famous case study in Psychology - which states that: as a person observes a long sequence of heads in a coin flipping process, he believes tails becomes more likely on the next flip. This behavior is also known as Law of Maturity. Previous statistical studies proposed modeling such situation under the bayesian perspective. In them there is the assumption of exchangeability and, as a result, maturity is an inadmissible behavior. In this study, a new model in which the Gambler's belief need not be a fallacy is presented. This one is the usual finite population model and, therefore, only operationally meaningful quantities (operational parameters) are involved. We define a class of distributions for the operational parameter which is named as tighter than the Binomial. Maturity is a consequence of the belief in the prior distributions of this class. Furthermore, a subclass which refers to the distributions that are second-order tighter than the Binomial is presented. For prior distributions of this subclass the predictive failure rate is increasing, which can be interpreted as the most extreme result of maturity. The results of this study may contribute on the judgment of how reasonable the assumption of infinite exchangeability is relative to typical human perception.

Nick Chater
University College London

Is the brain a Bayesian?

The brain and cognitive sciences have been swept by a Bayesian revolution. But people are notoriously poor at reasoning about probability. So is the brain really a Bayesian, or not? This talk considers recent experimental and theoretical work attempting to provide an answer.

Pawel Chwalinski
Middlesex University

Pseudo Bayesian classifier use towards network intrusion detection

Artificial Immune System is one of the most challenging and appealing approach regarding network intrusion detection and prevention. As human immune system is not fully explicable, not all mechanism can be used as immune system metaphors. Therefore, the research community have been actively interested in partial solutions such as Negative Selection Algorithm or Danger Theory strategy. Danger Theory is based on the idea of system toleration towards foreign entities unless they are not recognized as danger ones. Danger alerts are sent out while one of the body cell transmits a threat signal. The efficient body cells set has been in the interest of the academic researches. The set can be perceived as a contingency table consisting of network attributes (columns) and rows are thought of as classes of behavior relevant for a normal network state or attacks suspicion. The process of data collection and analysis is an adaptive statistical approach. The pseudo Bayesian classifier will be used for data classification. The improvement of Danger Theory will be the efficient body cell representation together with the analysis process which lead to more proficient intrusion detection.

Roger Cooke
Delft University of Technology and *Resources for the Future, Washington DC*

Expert judgment, then and now and then again
Results of a recent review of 63 applications of the classical model for expert judgment will be presented, together with insights, lessons learned and war stories.

Alireza Daneshkhah
University of Strathclyde

On the Robustness of Bayesian Networks to Learning from Non-conjugate Sampling
Recent results concerning the instability of Bayes Factor search over Bayesian Networks (BN's) lead us to ask whether learning the parameters of a selected BN might also depend heavily on the often rather arbitrary choice of prior density. Robustness of inferences to misspecification of the prior density would at least ensure that a selected candidate model would give similar predictions of future data points given somewhat different priors and a given large training data set. In this paper we derive new explicit total variation bounds on the calculated posterior density as the function of the closeness of the genuine prior to the approximating one used and certain summary statistics of the calculated posterior density. We show that the approximating posterior density often converges to the genuine one as the number of sample point increases and our bounds allow us to identify when the posterior approximation might not. To prove our general results we needed to develop a new family of distance measures called local DeRobertis distances. These provide coarse non-parametric neighbourhoods and allowed us to derive elegant explicit posterior bounds in total variation. The bounds can be routinely calculated for BNs even when the sample has systematically missing observations and no conjugate analyses are possible
[Joint work with Jim Smith]

Radboud Duintjer Tebbens
Delft University of Technology

Practical and technical considerations of the implementation in R of Cooke's classical model for combining subjective assessments.
Cooke's "classical model" provides a methodology for combining subjective assessments from experts into an aggregate distribution that represents a rational consensus of the uncertainty among the experts. Since appropriately measuring expert performance and combining their assessments involves various computational steps, several implementations of the method have been developed since its inception, including the Excalibur program, which was specifically developed for this purpose and is freely available, and implementations in Fortran and Microsoft Excel. Recently, the method has also been implemented and made available in the popular open-source statistics package R. We analyzed the implementation in R and considered its use in practice. While the availability of the "classical model" in R facilitates processing of the input and output of the model using other functionality offered in R, potential users should be aware of a number of inaccuracies in the implementation in R. We point out these technical inaccuracies and discuss practical considerations of the implementation of the "classical model" in R.

Guy Freeman
University of Warwick

Using dynamic staged trees for discrete time series data: robust prediction, model selection and causal analysis

A new graphical model is proposed for discrete-valued discrete-time data. We define the dynamic staged tree and implement a one-step ahead prediction algorithm using multi-process modelling and the power steady model that is robust yet also easy to interpret. We also demonstrate how to analyse causal hypotheses on this model class. We illustrate our techniques with a real educational example.

Simon French
Manchester Business School

Combining Expert Judgement – A review for decision makers

In a paper written over twenty years ago, I distinguished three contexts in which one might wish to combine expert judgements of uncertainty: the expert problem, the group decision problem and the text-book problem. Over the intervening years much has been written on the first two, which have the focus of a single decision context, but little on the third, though the closely related field of meta-analysis has developed considerably. With many developments in internet technology, particularly in relation to interactivity and communication, the text-book problem is gaining in importance since data and expert judgements can be made available over the web to be used by many different individuals to shape their own beliefs in many different contexts. Moreover, applications such as web-based decision support, e participation and e democracy are making algorithmic 'solutions' to the group decision problem attractive, but we know that such solutions are, at best, rare and, at worst, illusory.

In this paper I survey developments since my earlier paper, and then I turn to how expert judgement might be used within web-based group decision support, as well as in e participation and e democracy contexts. The latter points to a growing importance of the textbook problem and suggests that Cooke's principles for scientific reporting of expert judgement studies may need enhancing for such studies to be used by a wider audience.

Paul Garthwaite
Open University

Prior distribution elicitation for generalized linear and piecewise-linear models

This talk focuses on an elicitation method for quantifying subjective opinion about generalized linear models. Opinion between a continuous predictor variable and the dependent variable is modeled by a piecewise-linear function, giving a flexible model that can represent a wide variety of opinion. To quantify his or her opinions, the expert performs assessment tasks that involve drawing graphs and bar-charts to specify medians and other quantiles. Opinion about the regression coefficients is represented by a multivariate normal distribution whose parameters are determined from the assessments. The method requires interactive graphics and has been implemented in at least three different forms, including one by Kynn that is available as stand-alone software or as a WinBUGS add-in. The assessment tasks that the expert performs will be described and some experiments and other work where the method has been used will be discussed. Certain limitations of the model used to represent opinion will be noted and strategies to remove these limitations suggested. A new implementation of the method has recently been produced and this will be available for demonstration outside the talk.

[Joint work with Fadlalla Elfadaly]

Leanna House
Virginia Tech

An Application of Reification to a Rainfall-Runoff Computer Model

Deterministic computer models or simulators are used regularly to assist researchers in understanding the behavior of complex physical systems when real-world observations are limited. However, simulators are often imperfect representations of physical systems and introduce a layer of uncertainty into model-based inferences that is hard to quantify. To formalize the use of expert judgement in assessing simulator uncertainty, Goldstein and Rougier (2008) propose a method called reification that decomposes the discrepancy between simulator predicitions and reality by an improved, hypothetical computer model known as a ``reified'' model. One criticism of reification is that validation is, at best, challenging; only expert critiques can the validate the subjective judgements used to specify a reified model. For this paper, we created an artificial case study based on a well-used hydrological computer model and real data that allows us to validate reification using both data and expert judgment.

Silvia Liverani
University of Bristol

Guided Search for Rhythmic Regulation

An increasing number of microarray experiments produce time series of expression levels for many genes. Some recent clustering algorithms respect the time ordering of the data and are, importantly, extremely fast. The focus of this presentation is the development of such an algorithm on a microarray data set consisting of thousands of genes measured at several time points. Circadian rhythms control the timing of various physiological and metabolic processes and are regulated by genes acting in feedback loops.

It has been shown (Liverani et al., Bayesian Analysis, 2009) that in such a context, and in a Bayesian framework, it is possible to develop a cluster algorithm that is guided by cluster signatures. These signatures identify the genes, and therefore the clusters, of scientific interest and localising the search of the partition space. Then the clustering algorithm becomes orders of magnitude faster, not wasting time on parts of the partition space of no interest to the scientist. The aim of the presentation is to improve the algorithm by Liverani et al. (2009) by weighting the interesting genes rather than using a simple indicator discriminant. We propose to do this by approximating using the continuous score on the signatures obtained as a function of the means of the parameters.

Michail Papathomas
Imperial College London

Correlated Binary Variables and Multi-level Expert Probability Assessments

We present a fully parametric approach for updating beliefs regarding correlated binary variables, after marginal probability assessments based on information of varying quality are provided by an expert. The approach taken is motivated and illustrated through a practical application. This methodology (Papathomas, 2008) offers new insight into the parameters that control the dependence of the binary variables, and the relation of these parameters to the joint density of the probability assessments. A comprehensible elicitation procedure for the model parameters is put forward. Finally, we discuss a recent elicitation exercise which provided additional insight into the challenges related to this problem.

Papathomas, M. (2008): Correlated binary variables and multi-level probability assessments. Scandinavian Journal of Statistics. 35, 169-185

Artemis Parvizi
Middlesex University

Ontologies and Bayesian approach in threat detection

Threat detection is one of the major issues for bodies in charge of emergencies (e.g. fire brigades, ambulance services, police, etc). In this work, we will tackle this problem with an ontological approach. There are two ontologies: one represents the default knowledge; the second refines the first ontology. These ontologies will be derived from data, and the training procedure is important for its implementation. One problem that arises in this task is to determine what variables should contribute to which of the two ontologies. We propose to use Bayesian networks to derive the optimal data and information structures, and this way to facilitate the design of our ontologies.

Lawrence Phillips
London School of Economics

Approving New Drugs: How can Bayesian Decision Theory help?

European regulatory agencies that approve new drugs receive from pharmaceutical companies large dossiers of information which must be translated into a simple decision: approve or reject. This decision is the result of extensive discussions in meetings that rely solely on human judgment to piece together the collective impact of the data. No models are used to compare the benefits against the risks. My talk will support several ways that Bayesian Decision Theory (BDT) could support this decision-making process, and I will invite participants to discuss several questions relevant to the application of BDT.

Jesus Rios
Manchester Business School

Adversarial Risk Analysis for Counterterrorism Modelling.

Recent large scale terrorist attacks have raised interest in counterterrorism modelling. A unifying theme in that research is the need to develop methods for the analysis of decisions when there are intelligent adversaries ready to increase our risks. Most of the approaches have a clear game theoretic flavour, although there are also some decision analytic based approaches. We have recently introduced a framework for adversarial risk analysis, aimed at dealing with problems with intelligent opponents and uncertain outcomes. We explore here how adversarial risk analysis may cope with three main stylised counterterrorism models: sequential defend-attack, simultaneous defend-attack and defend-attack defend models.
[Joint work with David Rios Insua, Spanish Royal Academy of Sciences]

Jonathan Rougier
University of Bristol

Assessing and quantifying uncertainty in natural hazards

In natural hazards (volcanoes, earthquakes, floods etc) it is useful for modelling purposes to make a distinction between aleatory and epistemic uncertainty, where the former represents the inherent or natural uncertainty of the hazard, and the latter represents everything else. Natural hazards scientists are often reluctant to quantify epistemic uncertainty with probability, due in a large part to its subjective nature. But this challenge should be weighed against the additional problems that non-quantified uncertainty create for the risk manager and the policymaker. This talk explores these issues in the light of the recent NERC scoping study on natural hazards uncertainty and risk.

Teddy Seidenfeld
Carnegie Mellon University

Forecasting with Imprecise Probabilities [IP] – some preliminary findings

De Finetti's theory of coherent 2-sided previsions serves as the basis for numerous Imprecise Probability [**IP**] generalizations. In its original form, de Finetti's 2-person, zero-sum prevision-game begins with a class of bounded random variables $\mathcal{X}$ measurable with respect to some common space $\{\Omega, \mathcal{B}\}$. One player, the *bookie*, is required to post a "fair" (i.e., 2-sided) prevision $\boldsymbol{P}[X]$ for each $X \in \mathcal{X}$. The bookie's opponent, the *gambler*, may choose finitely many non-zero real numbers$\{\alpha_i\}$ where, when the state $\omega \in \Omega$ obtains, the payoff to the bookie is $\Sigma_i\, \alpha_i(X_i(\omega) - \boldsymbol{P}[X_i])$, and the opposite payoff, $-\Sigma_i\, \alpha_i(X_i(\omega) - \boldsymbol{P}[X_i])$, is for the gambler. That is, the bookie is obliged either to buy (when $\alpha > 0$), or to sell (when $\alpha < 0$) $|\alpha|$-many units of X at the price, $\boldsymbol{P}(X)$. Hence, the previsions are described as being *2-sided* or *fair*.

The bookie's previsions are *incoherent* if the gambler has a finite strategy that insures a uniformly negative payoff for the bookie, i.e., if there exist a finite set $\{\alpha_i\}$ and $\varepsilon > 0$ such that, for each $\omega \in \Omega$, $\Sigma_i\, \alpha_i(X_i(\omega) - \boldsymbol{P}[X_i]) < -\varepsilon$. Otherwise, the bookie's previsions are *coherent*. De Finetti's *Fundamental Theorem of Previsions* insures that a bookie's previsions are coherent if and only if there exists a finitely additive probability P that determines the expected values for each $X \in \mathcal{X}$, and these expected values are the bookie's previsions: $\mathbf{E}_P[X] = \mathbf{P}[X]$. This result extends to conditional expectations using called-off previsions. The bookie's called-off prevision for X given event F has payoff $\boldsymbol{F}\alpha(X(\omega) - \boldsymbol{P}_F[X])$, so that the transaction is called-off in case the event F does not obtain, when its indicator satisfies $\boldsymbol{F} = 0$.

One, well-studied approach to an **IP**-theory is to modify de Finetti's game. For each $X \in \mathcal{X}$ the bookie fixes a pair of 1-sided previsions, rather than a single 2-sided prevision. For each X, the bookie announces one rate $\underline{\boldsymbol{P}}[X]$ as a buying price, and a (possibly) different rate $\overline{\boldsymbol{P}}\,[X]$ as a selling price. Likewise, one defines a pair of 1-sided called-off previsions for X given event F. A modified **IP**-Fundamental Theorem obtains: A bookie's 1-sided previsions are coherent if and only if there is a maximal, non-empty (convex) set of finitely additive probabilities $\mathcal{P}$ where $\underline{\boldsymbol{P}}(X) = \text{infemum}_{P \in \mathcal{P}}\, \mathbf{E}_P[X]$ and $\overline{\boldsymbol{P}}\,[X] = \text{supremum}_{P \in \mathcal{P}}\, \mathbf{E}_P[X]$.

De Finetti was concerned with strategic aspects of his prevision game. A bookie might take advantage of knowledge about the gambler to **announce** previsions different from her/his own degrees of belief. For instance, though the bookie's fair price for the (indicator function $\boldsymbol{G}$ of the) event G is .50, but being confident that the gambler's fair price for $\boldsymbol{G}$ is .75, the bookie might announce a prevision $\boldsymbol{P}[\boldsymbol{G}] = .70$, anticipating that the gambler will find it profitable to buy units of $\boldsymbol{G}$ from him/her at this inflated price.

In order to mitigate such strategic aspects of his prevision-game, de Finetti used probabilistic forecasting subject to Brier score as an alternative framework for introducing a second but equivalent criterion of coherence. Consider probabilistic forecasting of events. Each $X \in \mathcal{X}$ is an indicator function. The bookie's previsions now are probabilistic forecasts subject to Brier score: squared-error loss. The penalty for a forecast $\mathrm{P}[\boldsymbol{G}]$ when the state $\omega \in \Omega$ obtains is $(\boldsymbol{G}(\omega) - \mathrm{P}[\boldsymbol{G}])^2$. The score for the conditional forecast $\boldsymbol{P}_F[\boldsymbol{G}]$, called-off if event F fails, is $\boldsymbol{F}(\omega)(\boldsymbol{G}(\omega) - \mathrm{P}[\boldsymbol{G}])^2$. The score from a finite set of forecasts is the sum of the separate forecasts. A set of forecasts, $\{\mathrm{P}[X]: X \in \mathcal{X}\}$, is coherent in the second sense if and only if there is no rival set of forecasts $\{\mathrm{P}'[X]: X \in \mathcal{X}\}$ where, for some finite subset of $\mathcal{X}$, the score for the rival P′-forecasts uniformly dominates the score for the matching P-forecasts. De Finetti established equivalence between these two senses of coherence. The second sense of coherence, operationalized with Brier score, does not involve strategic forecasting, in contrast with the opportunity for strategic action in the 2-person previsions-game.

In ongoing work, we investigate a modification of de Finetti's second criterion so that it serves as a basis for **IP** theory. Analogous to the use of 1-sided previsions, we distinguish a *lower forecast*, used to assess the penalty score when the event fails, from an *upper forecast*, used to assess the penalty

score when the event obtains. The issues we address with this generalization go beyond de Finetti's motivation to use the second criterion of coherence in order to avoid the threat of strategic behavior.
[Joint work with M.J.Schervish, and J.B.Kadane]

Glenn Shafer
Rutgers

Game-theoretic probability and its applications
The game-theoretic framework, introduced by Vovk and myself in 2001 (www.probabilityandfinance.com), uses game theory instead of measure theory as a mathematical framework for probability. Classical theorems are proven by betting strategies that multiply a player's stake by a large factor if the theorem's prediction fails. In this talk, I will discuss the basic idea of game-theoretic probability and two important applications: (1) defensive forecasting, (2) the game-theoretic interpretation of Bayesian and Dempster-Shafer inference

Richard Wilkinson
University of Nottingham

Utilising subjective beliefs about model error in the calibration of stochastic computer experiments
Models of physical systems are rarely perfect representations of the system, yet this is often ignored and inference done instead under the assumption that the model is perfect. Modellers often have subjective beliefs about the quality of their model, and we can try to elicit these beliefs to move from making inference about the model, to making inferences about the system. Approximate Bayesian computation (ABC) methods are a class of approximate algorithms for parameter estimation using only simulations from a computer model, and are increasingly being used in many different application areas. The quality of the approximation can be considered in terms of performing inference in the presence of added noise. We can control the distribution of this noise and use it to represent model error and/or measurement error. By doing this we can hope to move from making approximate statements about the model, to hopefully making more accurate statements about the system.

Robert Winkler
Duke University

Probability Forecasting, Probability Evaluation, and Scoring Rules: Expanding the Toolbox
Bayesian methods are inherently probabilistic, with uncertainty explicitly represented in terms of probabilities. In the subjective Bayesian framework, these probabilities are ultimately subjective, whether they represent judgmental forecasts or forecasts based on models chosen by an analyst. This raises questions about the incentives for "good" probability forecasts and for the evaluation of probability forecasts in light of the events that occur. Scoring rules were developed to provide both ex ante incentives for careful and truthful probability forecasts and ex post evaluation measures that investigate characteristics of probability forecasts and can identify "better" probability forecasters. This talk focuses on recent developments in scoring rules that enrich the set of available rules and raise some important probability forecasting and evaluation issues deserving of further attention.

Stan Yip
University of Exeter

Dynamic Factor Models for Attributing Sources of Uncertainty in Global Climate Predictions

Quantification of various sources of uncertainty to the total uncertainty in global climate predictions is essential for environment policy decision. In this paper, a general review on the partitioning sources of uncertainty with emphasis on the interpretability of the scientific interest is presented. Motivated by the problem of attributing uncertainty related to the deviation of climate models, emission scenario and natural variability, a Bayesian dynamic factor model is proposed for a better understanding on the structure of the sources of uncertainty. An example on 15 model simulation runs for 3 scenarios of global annual mean temperature is also given.