Model Selection And Model Averaging

Variable selection in linear regression models is an important aspect of many scientific analyses. We review several frequentist model selection techniques in the introductory chapter. Model uncertainty is one of the serious issues related to the model selection problem. One way this issue can be resolved is by using a Bayesian technique called Bayesian model averaging BMA. In Chapter 2, we discuss BMA techniques and illustrate the ideas with examples. An often used BMA approach to model selection is based on the so-called highest posterior probability model. In Chapter 3 we discuss several asymptotic properties of this model selection technique. Under a spike and slab hierarchy we find that the highest posterior model is total risk consistent for model selection, but that it also possesses some curious properties. Most important of these is a marked underfitting in finite samples, a phenomenon well noted in the literature for Bayesian Information Criterion BIC related procedures, but not often associated with highest posterior model selection. We employ a rescaling of the hierarchy and show the resulting rescaled spike and slab models mitigate the effects of underfitting due a perfect cancelation of a BIC like penalty term. By drawing upon an equivalence between the highest posterior model and the median model we consider the issue of how to calibrate rescaled spike and slab models by looking at their posterior inclusion probabilities. In Chapter 4 we describe a new spike and slab model for model space exploration and variable selection in linear regression models. Several theoretical features are In Chapter 4 we describe a new spike and slab model for model space exploration and variable selection in linear regression models. Several theoretical features are discussed to motivate the approach. An R package model Sampler has been developed and applications are presented. In Chapter 5 we present a more stable variable selection technique. We also discuss the issue of model selection uncertainty. Numerical examples are provided.

Along with many practical applications, Bayesian Model Selection and Statistical Modeling presents an array of Bayesian inference and model selection procedures. It thoroughly explains the concepts, illustrates the derivations of various Bayesian model selection criteria through examples, and provides R code for implementation. The author shows how to implement a variety of Bayesian inference using R and sampling methods, such as Markov chain Monte Carlo. He covers the different types of simulation-based Bayesian model selection criteria, including the numerical calculation of Bayes factors, the Bayesian predictive information criterion, and the deviance information criterion. He also provides a theoretical basis for the analysis of these criteria. In addition, the author discusses how Bayesian model averaging can simultaneously treat both model and parameter uncertainties. Selecting and constructing the appropriate statistical model significantly affect the quality of results in decision making, forecasting, stochastic structure explorations, and other problems. Helping you choose the right Bayesian model, this book focuses on the framework for Bayesian model selection and includes practical examples of model selection criteria.

Classification and regression problems characterized by the number (p) of predictor variables being relatively large to the sample size (n), called 'the large p small n problem', are common in educational sciences. Variable selection methods can resolve the problem by reducing variable dimensionality while maintaining prediction accuracy. However, traditional statistical approaches, such as stepwise regression models, cannot deal with the large p small n problem effectively. In this dissertation, I introduce variable importance measures (VIMs) from decision tree models to educational research to select a parsimonious classification model as well as evaluate their properties under different conditions. I also propose new VIMs based on Bayesian model averaging (BMA) and a hybrid approach of random forests and BMA (RF-BMA). In addition, I propose a cross-validated permutation (CV-permutation) threshold for random forest VIMs to identify informative variables. Using classification models, a series of simulation studies is conducted with four simulation factors: 10 VIM methods, five data models, two numbers of variables, and two samples sizes. Each combination of the four factors was replicated 1,000 times. In addition to the simulation studies, VIMs were applied to an education longitudinal study to select influential predictor variables on the prediction of six-year college graduation. As evaluation measures of VIMs, effectiveness, rank, and Brier score were used. This dissertation finds that random forests VIMs with the CV-permutation threshold performed better in preserving prediction accuracy than the other VIMs under the large p small n condition. Conversely, BMA outperformed tree-based VIMs with regard to the effectiveness of variable selection under the small p large n condition. RF-BMA performed as well as tree VIMs not only in variable selection but also in prediction accuracy under the large p condition. When $p > 30$, RF-BMA outperformed BMA in variable selection and prediction accuracy. Therefore, RF-BMA can be an attractive alternative VIM to BMA as well as to tree-based VIMs. The case study results showed that most tree VIMs selected the top five cognitive measures. The VIMs from BMA and RF-BMA selected additional variables besides the cognitive measures.

Three Essays on Model Selection in Time Series Econometrics

Bayesian model averaging, learning and model selection

An Asymptotic Theory for Model Selection Inference in General Semiparametric Problems

Bayesian Methods for Ecology

Statistical Postprocessing of Ensemble Forecasts

This paper extends the Bayesian Model Averaging framework to panel data models where the lagged dependent variable as well as endogenous variables appear as regressors. We propose a Limited Information Bayesian Model Averaging (LIBMA) methodology and then test it using simulated data. Simulation results suggest that asymptotically our methodology performs well both in Bayesian model averaging and selection. In particular, LIBMA recovers the data generating process well, with high posterior inclusion probabilities for all the relevant regressors, and parameter estimates very close to their true values. These findings suggest that our methodology is well suited for inference in short dynamic panel data models with endogenous regressors in the context of model uncertainty. We illustrate the use of LIBMA in an application to the estimation of a dynamic gravity model for bilateral trade.

Variable selection methods and model selection approaches are valuable statistical tools, which are indispensable for almost any statistical modeling question. This review considers first the use of information criteria for model selection. Such criteria provide an ordering of the considered models where the 'best' model is selected. Different modeling goals, though, might require different criteria to be used. Next, the effect of including a penalty in the estimation process is discussed. Also nonparametric estimation contains several aspects of model choice, such as the choice of the estimator to use and the selection of tuning parameters. Fourth, model averaging approaches are reviewed where estimators from different models are weighted to provide one final estimator. There are several ways to choose the weights, most of them result in data-driven, hence random, weights. Challenges for inference after model selection and inference for model averaged estimators are discussed.

Abstract: A Bayes factor between two models can be greatly affected by the prior distributions on the model parameters. When prior information is weak, very dispersed proper prior distributions are often used. This is known to create a problem for the Bayes factor when competing models that differ in dimension, and it is of even greater concern when one of the models is of infinite dimension. Therefore, we propose an innovative criterion called the calibrated Bayes factor, which uses training samples to calibrate the prior distributions so that they achieve a reasonable level of "information". The calibrated

Bayes factor is then computed as the Bayes factor over the remaining data. The level of "information" is tied to the concentration of training-updated prior distributions, which is carefully evaluated by monitoring the distribution of symmetrized Kullback-Leibler divergence between two likelihood functions drawn independently from the training-updated prior distribution. Monte Carlo Markov chain algorithms are widely used in this research to generate parameter draws from training-updated prior distributions. Subsampling is applied to reduce dependence among parameter draws.

Bayesian Model Selection and Statistical Modeling

Bayesian Theory and Applications

Calibrated Bayes Factor and Bayesian Model Averaging

Regression and Time Series Model Selection

Variable Importance Measures for Parsimonious Model Selection

This brilliantly structured and comprehensive volume provides exhaustive explanations of the concepts and philosophy of statistical modeling, together with a wide range of practical and numerical examples.

There is a rich history of work on model selection and averaging in the statistics literature. The Bayesian paradigm provides an approach to model selection which successfully overcomes the drawbacks for which frequentist hypothesis testing has been criticized. Most commonly, Bayesian model selection methods are based on the Bayes factor. Additionally, the Bayes factor has applications outside the realm of model selection, such as model averaging. In a formal sense, as a supplement to the prior odds, the Bayes factor produces the posterior odds for a pair of models. These posterior odds can be translated to posterior probabilities and yields a full posterior distribution that assigns a probability to each model as well as a distribution over the parameters for each model. Then the Bayesian model averaging provides better prediction by making inferences based on a weighted average over all of the models considered.

The interest in using Bayesian methods in ecology is increasing, however many ecologists have difficulty with conducting the required analyses. McCarthy bridges that gap, using a clear and accessible style. The text also incorporates case studies to demonstrate mark-recapture analysis, development of population models and the use of subjective judgement. The advantages of Bayesian methods, are also described here, for example, the incorporation of any relevant prior information and the ability to assess the evidence in favour of competing hypotheses. Free software is available as well as an accompanying web-site containing the data files and WinBUGS codes. Bayesian Methods for Ecology will appeal to academic researchers, upper undergraduate and graduate students of Ecology.

Least Squares Model Averaging by Prediction Criterion

Bayesian Data Analysis, Third Edition

Model Averaging

Estimating and Correcting the Effects of Model Selection Uncertainty

Outlines and Highlights for Model Selection and Model Averaging by Gerda Claeskens, Isbn

This book provides a concise and accessible overview of model averaging, with a focus on applications. Model averaging is a common means of allowing for model uncertainty when analysing data, and has been used in a wide range of application areas, such as ecology, econometrics, meteorology and pharmacology. The book presents an overview of the methods developed in this area, illustrating many of them with examples from the life sciences involving real-world data. It also includes an extensive list of references and suggestions for further research. Further, it clearly demonstrates the links between the methods developed in statistics, econometrics and machine learning, as well as the connection between the Bayesian and frequentist approaches to model averaging. The book appeals to statisticians and scientists interested in what methods are available, how they differ and what is known about their properties. It is assumed that readers are familiar with the basic concepts of statistical theory and modelling, including probability, likelihood and generalized linear models.

We examine house price forecastability across the 50 states using Dynamic Model Averaging and Dynamic Model Selection, which allow for model change and parameter shifts. By allowing the entire forecasting model to change over time and across locations, the forecasting accuracy improves substantially. The states in which housing markets have been the most volatile are the states where model change and parameter shifts have been the most needed.

Episode Treatment Groups (ETGs) classify related services into medically relevant and distinct units describing an episode of care. Proper model selection for those ETG based costs is essential to adequately price and manage health insurance risks. The optimal loss model (or model probabilities) can vary depending on the disease. We compare four potential models (lognormal, gamma, log-skew-t, and Lomax) using four different metrics (AIC and BIC weights, Random Forest feature classification, and Bayesian model averaging) on 320 episode treatment groups. Using the data from a major health insurer, which consists of more than 33 million observations from 9 million claimants, we compare the various methods on both speed and precision, and also examine the wide range of selected models for the different ETGs. Several case studies are provided for illustration. It is found that Random Forest feature selection is computationally efficient and sufficiently accurate, hence being preferred in this large data set. When feasible (on smaller data sets), Bayesian model averaging is preferred because of the posterior model probabilities.

Data Segmentation and Model Selection for Computer Vision

(comparison of Tree Models and Bayesian Model Averaging with Application to Prediction of College Graduation)

Focused Information Criterion and Model Averaging for Large Panels with a Multifactor Error Structure

Model Selection Versus Model Averaging in Dose Finding Studies

Model Selection and Inference

Statistical Postprocessing of Ensemble Forecasts brings together chapters contributed by international subject-matter experts describing the current state of the art in the statistical postprocessing of ensemble forecasts. The book illustrates the use of these methods in several important applications including weather, hydrological and climate forecasts, and renewable energy forecasting. After an introductory section on ensemble forecasts and prediction systems, the second section of the book is devoted to exposition of the methods available for statistical postprocessing of ensemble forecasts: univariate and multivariate ensemble postprocessing are first reviewed by Wilks (Chapters 3), then Schefzik and Möller (Chapter 4), and the more specialized perspective necessary for postprocessing forecasts for extremes is presented by Friederichs, Wahl, and Buschow (Chapter 5). The second section concludes with a discussion of forecast verification methods devised specifically for evaluation of ensemble forecasts (Chapter 6 by Thorarinsdottir and Schuhen). The third section of this book is devoted to applications of ensemble postprocessing. Practical aspects of ensemble postprocessing are first detailed in Chapter 7 (Hamill), including an extended and illustrative case study. Chapters 8 (Hemri), 9 (Pinson and Messner), and 10 (Van Schaeybroeck and Vannitsem) discuss ensemble postprocessing specifically for hydrological applications, postprocessing in support of renewable energy applications, and postprocessing of long-range forecasts from months to decades. Finally, Chapter 11 (Messner) provides a guide to the ensemble-postprocessing software available in the R programming language, which should greatly help readers implement many of the ideas presented in this book. Edited by three experts with strong and complementary expertise in statistical postprocessing of ensemble forecasts, this book assesses the new and rapidly developing field of ensemble forecast postprocessing as an extension of the use of statistical corrections to traditional deterministic forecasts. Statistical Postprocessing of Ensemble Forecasts is an essential resource for researchers, operational practitioners, and students in weather, seasonal, and climate forecasting, as well as users of such forecasts in fields involving renewable energy, conventional energy, hydrology, environmental engineering, and agriculture. Consolidates, for the first time, the methodologies and applications of ensemble forecasts in one succinct place Provides real-world examples of methods used to formulate forecasts Presents the tools needed to make the best use of multiple model forecasts in a timely and efficient manner

Rainfall-Runoff Modelling: The Primer Second Edition focuses on predicting hydrographs using models based on data and on representations of hydrological process. Dealing with the history of the development of rainfall-runoff models, uncertainty in mode predictions, good and bad practice and ending with a look at how to predict future catchment hydrological responses this book provides an essential underpinning of rainfall-runoff modelling topics."--pub. desc.

Model Selection and Model Averaging

Model Selection and Multimodel Inference

Model Selection and Model Averaging in the Presence of Missing Values

Statistical Model Selection in Operational Risk

Bayesian Variable Selection and Model Averaging in High Dimensional Multinominal Nonparametric Regression

A Statistical Approach

Never HIGHLIGHT a Book Again! Virtually all of the testable terms, concepts, persons, places, and events from the textbook are included. Cram101 Just the FACTS101 studyguides give all of the outlines, highlights, notes, and quizzes for your textbook with optional online comprehensive practice tests. Only Cram101 is Textbook Specific. Accompanys: 9780521852258 .

First book to synthesize the research and practice from the active field of model selection.

Statisticians and applied scientists must often select a model to fit empirical data. This book discusses the philosophy and strategy of selecting such a model using the information theory approach pioneered by Hirotugu Akaike. This approach focuses critical attention on a priori modeling and the selection of a good approximating model that best represents the inference supported by the data. The book includes practical applications in biology and environmental science.

Model Averaging, Causal Graphs, and Structural Identification

Limited Information Bayesian Model Averaging for Dynamic Panels with An Application to a Trade Gravity Model

Model Selection

Bayesian Averaging, Prediction and Nonnested Model Selection

Forecasting House Prices in the 50 States Using Dynamic Model Averaging and Dynamic Model Selection

A unique and comprehensive text on the philosophy of model-based data analysis and strategy for the analysis of empirical data. The book introduces information theoretic approaches and focuses critical attention on a priori modeling and the selection of a good approximating model that best represents the inference supported by the data. It contains several new approaches to estimating model selection uncertainty and incorporating selection uncertainty into estimates of precision. An array of examples is given to illustrate various technical issues. The text has been written for biologists and statisticians using models for making inferences from empirical data.

"Bayesian variable selection has experienced substantial developments over the past 30 years with the proliferation of large data sets. Identifying relevant variables to include in a model allows simpler interpretation, avoids overfitting and multicollinearity, and can provide insights into the mechanisms underlying an observed phenomenon. Variable selection is especially important when the number of potential predictors is substantially larger than the sample size and sparsity can reasonably be assumed. The Handbook of Bayesian Variable Selection provides a comprehensive review of theoretical, methodological and computational aspects of Bayesian methods for variable selection. The topics covered include spike-and-slab priors, continuous shrinkage priors, Bayes factors, Bayesian model averaging, partitioning methods, as well as variable selection in decision trees and edge selection in graphical models. The handbook targets graduate students and established researchers who seek to understand the latest developments in the field. It also provides a valuable reference for all interested in applying existing methods and/or pursuing methodological extensions"--

While substantial amount of work in operational risk has been dedicated to fitting distributions to the loss data, comparatively little was written about model selection. Arguably, a good selection procedure is more important than fitting one since it can rectify a failure of fitting methodology by rejecting a bad model. This presentation applies the work of K.P. Burnham and D.R. Anderson, 2002 in the context of the operational risk capital modeling, including the use of information criteria and model averaging.

Model selection and model averaging

Model Selection and Model Averaging for Neural Networks

A Practical Information-Theoretic Approach

9780521852258

Calibrated Bayes Factors for Model Selection and Model Averaging

This paper studies the asymptotic relationship between Bayesian model averaging and post-selection frequentist predictors in both nested and nonnested models. We derive conditions under which their difference is of a smaller order of magnitude than the inverse of the square root of the sample size in large samples. This result depends crucially on the relation between posterior odds and frequentist model selection criteria. Weak conditions are given under which consistent model selection is feasible, regardless of whether models are nested or nonnested and regardless of whether models are correctly specified or not, in the sense that they select the best model with the least number of parameters with probability converging to 1. Under these conditions, Bayesian posterior odds and BICs are consistent for selecting among nested models, but are not consistent for selecting among nonnested models.

This volume guides the reader along a statistical journey that begins with the basic structure of Bayesian theory, and then provides details on most of the past and present advances in this field.

Now in its third edition, this classic book is widely considered the leading text on Bayesian methods, lauded for its accessible, practical approach to analyzing data and solving research problems. Bayesian Data Analysis, Third Edition continues to take an applied approach to analysis using up-to-date Bayesian methods. The authors—all leaders in the statistics community—introduce basic concepts from a data-analytic perspective before presenting advanced methods. Throughout the text, numerous worked examples drawn from real applications and research emphasize the use of Bayesian inference in practice. New to the Third Edition Four new chapters on nonparametric modeling Coverage of weakly informative priors and boundary-avoiding priors Updated discussion of cross-validation and predictive information criteria Improved convergence monitoring and effective sample size calculations for iterative simulation Presentations of Hamiltonian Monte Carlo, variational Bayes, and expectation propagation New and revised software code The book can be used in three different ways. For undergraduate students, it introduces Bayesian inference starting from first principles. For graduate students, the text presents effective current approaches to Bayesian modeling and computation in statistics and related fields. For researchers, it provides an assortment of Bayesian methods in applied statistics. Additional materials, including data sets used in the examples, solutions to selected exercises, and software instructions, are available on the book's web page.

Prediction and Variable Selection

Rainfall-Runoff Modelling

Statistical Model Choice

Handbook of Bayesian Variable Selection

This important book describes procedures for selecting a model from a large set of competing statistical models. It includes model selection techniques for univariate and multivariate regression models, univariate and multivariate autoregressive models, nonparametric (including wavelets) and semiparametric regression models, and quasi-likelihood and robust regression models. Information-based model selection criteria are discussed, and small sample and asymptotic properties are presented. The book also provides examples and large scale simulation studies comparing the performances of information-based model selection criteria, bootstrapping, and cross-validation selection methods over a wide range of models.

This paper considers model selection and model averaging in panel data models with a multifactor error structure. We investigate the limiting distribution of the common correlated effects estimator (Pesaran, 2006) in a local asymptotic framework and show that the trade-off between bias and variance remains in the asymptotic theory. We then propose a focused information criterion and a plug-in averaging estimator for large heterogeneous panels and examine their theoretical properties. The novel feature of the proposed method is that it aims to minimize the sample analog of the asymptotic mean squared error and can be applied to cases irrespective of whether the rank condition holds or not. Monte Carlo simulations show that both proposed selection and averaging methods generally achieve lower expected squared error than other methods. The proposed methods are applied to analyze the consumer response to gasoline taxes.

This edited volume explores several issues relating to parametric segmentation including robust operations, model selection

criteria and automatic model selection, plus 2D and 3D scene segmentation. Emphasis is placed on robust model selection with techniques such as robust Mallows Cp, least K-th order statistical model fitting (LKS), and robust regression receiving much attention. With contributions from leading researchers, this is a valuable resource for researchers and graduated students working in computer vision, pattern recognition, image processing and robotics.

Model Selection and Averaging of Health Costs in Episode Treatment Groups

Model Selection and Model Averaging

????????

The Primer

Information Criteria and Statistical Modeling

Most applied statistical analyses are carried out under model uncertainty, meaning that the model which generated the observations is unknown, and so the data are first used to select one of a set of plausible models by means of some selection criterion. Generally the data are then used to make inferences about some quantity of interest, ignoring model selection uncertainty, i.e. the fact that the selection step was carried out using the same data, and despite the known fact that this leads to invalid inferences. This thesis investigates several issues relating to this problem from both the Bayesian and the frequentist points of view, and offers new suggestions for dealing with it. We examine Bayesian model averaging (BMA) and point out that its frequentist performance is not always well-defined because, in some cases, it is unclear whether BMA methodology is truly Bayesian. We illustrate the point with a “fully Bayesian model averaging" that is applicable when the quantity of interest is parametric.