Lecture Notes in Statistics

Edited by D. Brillinger, S. Fienberg, J. Gani,
J. Hartigan, and K. Krickeberg

16

Specifying Statistical Models

From Parametric to Non-Parametric, Using Bayesian or Non-Bayesian Approaches

Edited by
J.P. Florens, M. Mouchart, J.P. Raoult,
L. Simar, and A.F.M. Smith

Springer-Verlag
New York Heidelberg Berlin

Lecture Notes in Statistics

Edited by D. Brillinger, S. Fienberg, J. Gani,
J. Hartigan, and K. Krickeberg

16

Specifying Statistical Models

From Parametric to Non-Parametric, Using Bayesian or Non-Bayesian Approaches

Edited by J.P. Florens,
M. Mouchart, J.P. Raoult,
L. Simar, and A.F.M. Smith

Springer-Verlag
New York Heidelberg Berlin

J.P. Florens
Université d'Aix-Marseille II
France

J.P. Raoult
Université de Rouen
France

A.F.M. Smith
University of Nottingham
United Kingdom

M. Mouchart
C.O.R.E.
Université Catholique de Louvain
Belgium

L. Simar
Facultés Universitaires Saint-Louis
Bruxelles
Belgium

AMS Classifications: 62-06, 62CXX, 62D05

Library of Congress Cataloging in Publication Data
Franco-Belgian Meeting of Statisticians (2nd: 1981:
Louvain-la-Neuve, Belgium)
Specifying statistical models (from parametric to
non-parametric, using Bayesian or non-Bayesian approaches)
(Lecture notes in statistics; 16)
1. Mathematical statistics—Congresses. I. Florens,
J.P. II. Title. III. Series: Lecture notes in statistics
(Springer-Verlag); 16.
QA276.A1F73 1981 519.5 82-19314

Proceedings of the Second Franco-Belgian Meeting of Statisticians held in Louvain-la-Neuve (Belgium) on October 15–16, 1981.

With 24 Illustrations

Printed and bound by R.R. Donnelley & Sons, Harrisonburg, VA.
Printed in the United States of America.

9 8 7 6 5 4 3 2 1

ISBN 0–387–90809-9 Springer-Verlag New York Heidelberg Berlin
ISBN 3–540–90809-9 Springer-Verlag Berlin Heidelberg New York

Lecture Notes in Statistics

Vol. 1: R. A. Fisher: An Appreciation. Edited by S. E. Fienberg and D. V. Hinkley. xi, 208 pages, 1980.

Vol. 2: Mathematical Statistics and Probability Theory. Proceedings 1978. Edited by W. Klonecki, A. Kozek, and J. Rosiński. xxiv, 373 pages, 1980.

Vol. 3: B. D. Spencer, Benefit-Cost Analysis of Data Used to Allocate Funds. viii, 296 pages, 1980.

Vol. 4: E. A. van Doorn, Stochastic Monotonicity and Queueing Applications of Birth-Death Processes. vi, 118 pages, 1981.

Vol. 5: T. Rolski, Stationary Random Processes Associated with Point Processes. vi, 139 pages, 1981.

Vol. 6: S. S. Gupta and D.-Y. Huang, Multiple Statistical Decision Theory: Recent Developments. viii, 104 pages, 1981.

Vol. 7: M. Akahira and K. Takeuchi, Asymptotic Efficiency of Statistical Estimators. viii, 242 pages, 1981.

Vol. 8: The First Pannonian Symposium on Mathematical Statistics. Edited by P. Révész, L. Schmetterer, and V. M. Zolotarev. vi, 308 pages, 1981.

Vol. 9: B. Jørgensen, Statistical Properties of the Generalized Inverse Gaussian Distribution. vi, 188 pages, 1981.

Vol. 10: A. A. McIntosh, Fitting Linear Models: An Application of Conjugate Gradient Algorithms. vi, 200 pages, 1982.

Vol. 11: D. F. Nicholls and B. G. Quinn, Random Coefficient Autoregressive Models: An Introduction. v, 154 pages, 1982.

Vol. 12: M. Jacobsen, Statistical Analysis of Counting Processes. vii, 226 pages, 1982.

Vol. 13: J. Pfanzagl (with the assistance of W. Wefelmeyer), Contributions to a General Asymptotic Statistical Theory. vii, 315 pages, 1982.

Vol. 14: GLIM 82: Proceedings of the International Conference on Generalised Linear Models. Edited by R. Gilchrist. v, 188 pages, 1982.

Vol. 15: K. R. W. Brewer and M. Hanif, Sampling with Unequal Probabilities. vii, 164 pages, 1983.

Vol. 16: Specifying Statistical Models: From Parametric to Non-Parametric, Using Bayesian or Non-Bayesian Approaches. Edited by J. P. Florens, M. Mouchart, J. P. Raoult, L. Simar, and A. F. M. Smith. xi, 204 pages, 1983.

PREFACE

During the last decades, the evolution of theoretical statistics has been marked by a considerable expansion of the number of mathematically and computationaly tractable models. Faced with this inflation, applied statisticians feel more and more uncomfortable : they are often hesitant about their traditional (typically parametric) assumptions, such as normal and i.i.d., ARMA forms for time-series, etc., but are at the same time afraid of venturing into the jungle of less familiar models. The problem of the justification for taking up one model rather than another one is thus a crucial one, and can take different forms.

(a) Specification : Do observations suggest the use of a different model from the one initially proposed (e.g. one which takes account of outliers), or do they render plausible a choice from among different proposed models (e.g. fixing or not the value of a certain parameter) ?

(b) Model Approximation : How is it possible to compute a "distance" between a given model and a less (or more) sophisticated one, and what is the technical meaning of such a "distance" ?

(c) Robustness : To what extent do the qualities of a procedure, well adapted to a "small" model, deteriorate when this model is replaced by a more general one ? This question can be considered not only, as usual, in a parametric framework (contamination) or in the extension from parametric to non parametric models but also, within a non parametric framework, for evaluating the "weight" of some technical hypothesis (such as markovicity, knowledge of the stationary measure of a process, properties of a regression function, etc.).

(d) Adaptivity : Once one decides to step outside a traditional framework, is it possible to adapt procedures in current use within such a framework so as to obtain tractable procedures (e.g. the choice of a tractable prior measure in a non parametric framework, the computation of an approximate bayesian estimator instead of an intractable exact one, the enlarging of the null hypothesis w.r.t. which a statistic is parameter free, etc.) ?

The 12 lectures collected in these Proceedings were all presented at the "Rencontre Franco-Belge de Statisticiens" to an audience of research workers in applied and theoretical statistics. Different classifications could be possible for these papers : according to the degree of generality of their subjects (papers [1] to [5] were specially considered as methodology papers, even if they contain some specific results), according to the type of statistical theory to which they refer (papers [1], [2], [6], [7] and [8] belong to Bayesian statistics), according to their probabilistic framework (papers [4], [5], [11] and, partially, [2] and [9] are devoted to statistics on random processes; paper [7] uses the classical tools of linear statistics).

The papers are printed in the order of presentation at the conference. Papers [1] to [5] were presented the first day, oriented toward a larger audience. Some of the other papers are more technical. In this introduction, we have classified then according to the four types of methodological considerations that are listed above.

(a) Specification

Bayesian Statistics provide a natural framework for specification problems (and Box and Tiao's work in 1973 is quite enlightning in this respect). Roughly speaking, a "huge" model can always be considered as an union of smaller ones, and supplied with a prior measure. The consideration of the a posteriori measure over different small models does not present any theoretical difficulty; however, such unions of models may present some intricacies of interpretation (e.g. is there a common meaning for parameters having the same name inside the different models ?) and some computational difficulties which are analyzed in [1] (L. Simar); a review of some classical examples in this domain is given in [2] (A.F.M. Smith), where new results are presented for the treatment of outliers (in the univariate case, the different small models among which one has to choose are characterized by the number of possible lower and upper outliers; a multivariate generalization is sketched).

(b) Model Approximation

In Bayesian statistics, there are different ways of defining a distance between a model and a submodel; if the submodel is obtained by reduction on one only of the two component spaces (parameter space and sample space) one gets a notion of approximate sufficiency; various such notions are studied in [6] (J.P. Florens); if the reduction is made on the parameter space, ε-sufficiency (with ε sufficiently small) can be a justification for adopting the reduced model instead of the complete one.

In a non-bayesian framework, one can extend a given model by supposing that the parameter θ (initially considered as deterministic) may be submitted to a random pertubation in the neighbourhood of a fixed (but unknown) value; for i.i.d. observations (in the initial model), A. Hillion computes in [10] how the distance (variation distance, or Hellinger distance) between probabilities (for an infinite sequence of observations) in the initial model or in the perturbed model are bounded according to the magnitude of the perturbation; convergence results follow.

(c) Robustness

Classical statistical analyses of stationary time series rely heavily on the Box Jenkins approach for ARMA models; the specification of the order of these models is often heuristically and technically very difficult; in such cases, it is natural to propose non parametric estimators, for instance for the conditional expectation of $g(x_{n+s})$, given $(x_1, \ldots, x_n)$; convergence properties of such kernel estimators (based on x_n and on all couples $(x_i, g(x_{i+s}))$ $(1 \leqslant i \leqslant n-s)$ are studied in [5] (D. Bosq). Although heuristically markovian, these estimators rely only on mixing properties of

the process and on hypothesis connecting the kernel and the stationary measure; in case this stationary measure is unknown, robustness properties (w.r.t. to bad choices of the kernel) are given in terms of the upper bound of the mean quadratic risk.

The same problem, for multivariate mixing processes, is studied in [11] (G. Collomb). In this paper, the estimators, called predictograms, are not based on general kernels, but on partitions of the space of observations $(x_1, \ldots, x_n)$ and make use of the k-uple $(x_{n-k+1}, \ldots, x_n)$ and of all the (k+1)-uples $(x_{i-k+1}, \ldots, x_i, g(x_{i+s}))$ $(k \leqslant i \leqslant n-s)$. Comparison of the speeds of convergence given in [5] and [11] can be useful for the choice of the number of successive observations used in the computations (i.e. typically, though the results do not rely directly on Markov properties, the order of markovicity assumed in the model).

In [4] (P. Doukhan), attention is focused on autoregressive Markov processes $(x_{n+1} = f(x_n) + \varepsilon_n)$; consistent kernel estimators of f are known, under hypothesis of regularity on f $(\|f\|_\infty + \|f'\|_\infty < \infty)$; their mean quadratic risk is of order $n^{-2/3}$, up to multiplicative constants in which both f and the variance σ^2 of the noise ε_n intervene; simulations are made, in [4], in order to study the deterioration of the qualities of these estimators for some "bad" functions of regression (i.e. not satisfying the hypothesis of the convergence theorem) and their sensitivity to the variations of σ^2 and to the shape of f (e.g. a periodic f, with its period small w.r.t. σ, constitutes a model which is indistinguishable from an i.i.d. model).

In the i.i.d. case, it is a classical problem to test (by means of "least favourable couples") a ball $B(P,r)$ against an other $B(P',r')$ (the radii being measured with the Hellinger distance). The error probabilities are known to decrease exponentially to 0; in [9] (L. Birge) explicit computations of the way these error probabilities increase when the "reality" (not necessarily i.d.) lies outside of one of these balls are given; for example, among n observations, all laws slightly outside $B(P,r)$, or a small number among the n observations distinctly outside $B(P,r)$. As a consequence of these computations, one gets properties of robustness w.r.t. the choice of radii; in other words one gets an answer to the question : what is the penalty if an attempt of robustification, leading to a model with the ball $B(P,r)$ instead of its center P, has not been strong enough ?

(d) Adaptivity

The classical robust tests between neighbourhoods of given probabilities, as seen in the last paragraph of (c) above, are no more tractable when the centers of these neighbourhoods do not correspond to i.i.d. models. In such cases (i. non i.d., or Markov) one has to use generalizations of the Hellinger distance; balls are not convex any more, and the "least favorable couple" argument does not hold. In the same paper [9], L. Birge provides adaptations of the classical tests of neighbourhoods, keeping the fundamental property of exponentially decreasing error probabilities.

In multivariate non parametric testing of hypothesis, it is sometimes proposed that test functions be constructed by means of parameter free statistics whose law is multidimensional uniform with respect to any probability in the null hypothesis (this is for instance quite usual for testing multidimensional normality). Such parameter-free statistics are constructed by using sufficient statistics whose laws are absolutely continuous. In [3] (J.P. Raoult, D. Criticou, D. Terzakis) this type of method is adapted to the non absolutely continuous case (tests are then randomized tests).

Bayesian non parametric theory, which had been neglected for a long time, became tractable essentially after the introduction (Ferguson, Doksum, 1973) of Dirichlet processes which allow easy computations of posterior probabilities on the set of all distribution functions on $\mathbb{R}$. In [8] (J.M. Rolin) attention is focused on the "general theory of processes" properties which make these computations feasible. This paper characterizes the Dirichlet process by stochastic independence properties rather than by distributional properties. Some extensions of these independence properties lead to consider neutral processes as a natural class of tractable prior probabilities.

A final adaptivity problem, in a Bayesian framework, is the one studied in [7] (M. Mouchart et L. Simar). Computation of the posterior expectation of the parameter, $E(\theta/x)$ is often difficult once one decides to step outside a traditional model. Least squares approximations within more general models are shown to provide tractable adaptation; the role of exchangeability conditions is studied in this respect.

Finally, paper [12] (J. Benasseni) is devoted to principal component analysis : once chosen weights for the observations (in order to define a distance between the variables), the usual procedures make use of the centering of the observations which is connected to those weights but, this connection does not exist any more if the computations are done by using weights and centering constant which are respectively the coefficients appearing in robust estimates of mean and variance. It is shown that however this procedure can still be considered as principal component analysis provided the distance between the variable is suitably defined.

The editors wish to thank the Fonds National de la Recherche Scientifique, the Center for Operations Research and Econometrics (C.O.R.E. - Université Catholique de Louvain) and the Séminaire de Mathématiques Appliquées aux Sciences Humaines (SMASH - Facultés Universitaires Saint-Louis - Bruxelles) for their support to this meeting.

The Editors,

Florens, J.-P.
Mouchart, M.
Raoult, J.-P.
Simar, L.
Smith, A.F.M.

LIST OF PARTICIPANTS

BALLET, M.F.
Boels et Bégault
Rue des Blancs Chevaux 7-301
1348 Louvain-la-Neuve

BENASSENI, J.
CRIG Montpellier
"Résidence les Rainettes" Bat. C
Rue de Las Sorbes, 1250
Montpellier - France

BENMANSOUR, D.
Université de Rouen
Cité Panorama, 112 Flaubert
76130 Mont-Saint-Signon

BERTRAND, D.
U.C.L.
Chemin Long, 213
1310 La Hulpe

BIRGE, L.
Université de Paris X, Nanterre
Boulevard Magenta, 93
75010 Paris - France

BIRONT, E.
U.C.L.
Chaussée de Boondael, 306
1050 Bruxelles

BOSQ, D.
Université de Lille I
Rue Dunois, 74
75013 Paris - France

CARLETTI, G.
IRSIA - Bureau de Biométrie
Rue de Roncia, 23
5800 Gembloux

COCCHI, D.
CORE - U.C.L.
Rue de la Neuville, 30
1348 Louvain-la-Neuve

COLLOMB, G.
U.P.S.
Rue de l'Homme Armé, 1
31000 Toulouse - France

DEHANDTSCHUTTER, M.
U.C.L.
Rue des Blancs Chevaux, 1/101
1348 Louvain-la-Neuve

de LEVAL, N.
Fac. de Médecine - U.C.L.
Rue des Cours, 13
5865 Walhain

DEPRINS, J.
U.L.B.
Stanley, 7
1980 Tervuren

DEPRINS-VANHECKE, D.
Facultés Universitaires S^{t} Louis
Avenue des Prisonniers de Guerre, 5
1490 Court-Saint-Etienne

DOUKHAN, P.
Université de Rouen
Rue de la Croix Nivert, 214
75015 Paris - France

DREZE, J.
CORE - U.C.L.
Voie du Roman Pays, 34
1348 Louvain-la-Neuve

EMBRECHTS P.
K.U.L.
Fonteinstraat, 58/2
3031 Oud-Heverlee

FLORENS, J.-P.
Université d'Aix Marseille
Le Corbusier, 229 -Bd Michelet
13008 Marseille - France

FOURDRINIER, D.
Université de Rouen
Rue de Beauvoisine, 12
76000 Rouen - France

GOVAERTS, B.
CORE - U.C.L.
Avenue M. César, 89
1970 Wezembeek

GRANCHER, G.
Rue Roger Gobbé, 354
76230 Bois-Guillaume - France

GUERIT, Y.
Crédit Communal de Belgique
Rue du Grand Marais, 11
7410 Mons

HALLIN, M.
U.L.B.
Avenue de Catus, 6
1310 La Hulpe

HILLION, A.
Université de Bretagne
Occidentale
Rue Joseph le Brix, 20
29200 Brest - France

HOUTMAN, A.
CORE - U.C.L.
Avenue Chataigniers, 3
1640 Rhode-Saint-Genèse

JANSSEN, J.
U.L.B.
Avenue J. Stobbaerts, 59
1030 Bruxelles

JOYEUX, R.
CORE - U.C.L. et Cornell Univ. (U.S.A.)
Rue des Pêcheries, 103 Boîte 50
1160 Bruxelles

KESTEMONT, M.-P.
F.U.S.L. et U.C.L.
Rue Klakkedelle, 68
1200 Bruxelles

KUPPER, J.-M.
Grand Monchaut, 74
7890 Ellezelles

LUBRANO, M.
CORE - U.C.L.
Rue du Sablon, 32/405
1348 Louvain-la-Neuve

MACHTELINCKX
U.L.B.
Avenue A. Giraud, 48
1030 Bruxelles

MAES, J.-M.
Avenue de Tervuren, 441
1150 Bruxelles

MALBECQ, W.
U.L.B.
Avenue Général Lebon, 113, Bte 7
1160 Bruxelles

MERCIER , F.
U.C.L.
Rue A. Delvaux, 17
6040 Charleroi

MIKHAEL, A.
Université de Rouen
Fac. des Sciences
Bâtiment de Mathématiques
76130 Mont-Saint-Aignan - France

MOUCHART, M.
CORE - U.C.L.
Rue Haute, 47
1348 Louvain-la-Neuve

ORSI, R.
CORE - U.C.L.
Rue de la Neuville, 30
1348 Louvain-la-Neuve

PARIS, J.
U.C.L.
Boulevard Paul Janson, 88
6000 Charleroi

RAOULT, J.-P.
Université de Rouen
Avenue Ganbetta, 15
92410 Ville d'Avray - France

REY, W.
M.B.L.E. - Bruxelles
Boulevard du Souverain, 138, Bte 19
1190 Bruxelles

ROLIN, J.-M.
U.C.L.
Rue Elizabeth, 1
5800 Gembloux

RUTGEERTS, A.-M.
ADPU - U.C.L.
Rue des Blancs Chevaux, 1/401
1348 Louvain-la-Neuve

SIMAR, L.
Facultés Universitaires S^{t} Louis
Avenue des Chênes, 16
5870 Mont-Saint-Guibert

SMITH, A.F.M.
University of Nottingham
Nether St., 41
Bekston, Nottingham, U.K.

SPIES, J.-M.
S.N.C.B.
Clos des Pommiers, 29
1310 La Hulpe

VAES, T.
IAG - U.C.L.
Rue de Parme, 45
1060 Bruxelles

VAN CUTSEM, B.
Univ. Scient. et Méd. de Grenoble
Avenue Louis Noiray Corenc, 7.
38700 La Tronche - France

VANDAUDENARD, R.
U.C.L.
Rue Brichaut, 16
1030 Bruxelles

CONTENTS

PROTECTING AGAINST GROSS ERRORS : THE AID OF BAYESIAN METHODS

by

Léopold Simar

Facultés Universitaires Saint-Louis, Bruxelles
and Center for Operations Research and Econometrics
Université Catholique de Louvain

Abstract

A statistical model is characterized by a family of probability distribution functions. All inferences are then conditional on the hypothesis formalised by this family.

The statistician often needs to protect himself against the consequences of a gross error relative to the basic hypothesis : either a specification error for the functionnal form of $p(x|\theta)$, or the treatment of outliers. It will be shown in this paper that the Bayesian approach offers a natural framework for treating this kind of problem. Different methods are presented : robustness analysis considering the sensitivity of inference to the model specification; and approximations to Bayesian solutions which are for a large class of models and sometimes preferable to the exact solutions valid only for a particular model.

Key-words : Bayesian sensitivity, Robustness analysis.

AMS/MOS : Primary 62A15, Secondary 62G35.

Acknowledgements

I would like to express my thanks to M. Mouchart with whom I have had numerous discussions on the subject of this paper and to J.-M. Rolin for his help in revising the paper. Errors and shortcomings are exclusively my responsibility.

1. INTRODUCTION

Many statistical procedures are based on statistical models which specify under which conditions the data are generated. Usually the assumption is made that the set of observations $x_1, \ldots, x_n$ is a set of (i) independent random variables (ii) identically distributed with common p.d.f. $p(x_i|\theta)$. Once this model is specified, the statistician tries to find optimal solutions to his problem (usually related to the inference on a set of parameters $\theta \in \Theta \subset \mathbf{R}^k$, characterizing the uncertainty about the model).

The procedure just described is not always easy to carry out. In fact when confronted with a set of data three attitudes are possible :

(1) The statistician may be a "pessimist" who does not belief in any particular model $p(x|\theta)$. In this case he must be satisfied with descriptive methods (like exploratory data analysis) without the possibility of inductive inference.

(2) The statistician may be an "optimist" who strongly believes in one model. In this case the analysis is straightforward and optimal solutions may often be easily obtained.

(3) The statistician may be "realist" : he would like to specify a particular model $p(x|\theta)$ in order to get operational results but he may have either some doubt about the validity of this hypothesis or some difficulty in choosing a particular parametric family.

Let us illustrate this kind of preoccupation with an example. Suppose that the parameter of interest is the "centre" of some population. In many situations, the statistician may argue that, due to a central limit effect, the data are generated by a normal p.d.f. In this case the problem is restricted to the problem of inference on μ, the mean of the population. But in some cases, he may have some doubt about these central limit effects and may suspect some skewness and/or some kurtosis or he may suspect that some observations are generated by other models (leading to the presence of outliers).

In this context three types of question may be raised in order to avoid gross errors in the prediction, or in the inference :

(i) Does the optimal solution, computed for an assumed model $p(x|\theta)$, still have "good" properties if the true model is a little different ?

(ii) Are the optimal solutions computed for other models near to the original one really substantially different ?

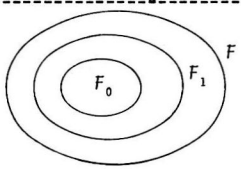
(iii) Is it possible to compute (exactly or approximatively) optimal solutions for a wider class of models based on very few assumptions ?

The first question is concerned with the sensitivity of a given criterion to the hypothesis (criterion robustness). In the second question, it is the sensitivity of the inference which is analysed (inference robustness). The last question may be viewed as a tentative first step towards the development of nonparametric methods (i.e. methods based on a very large parametric space).

The object of the paper is to provide some insight into the problems raised by the two latter questions. It will be shown that a sensitivity (robustness) analysis is natural in a Bayesian framework and that distribution-free methods (valid for a large class of models) can be obtained. Different approaches will be considered, which can be schematically presented as follows.

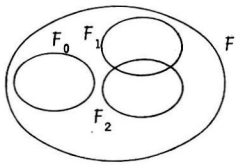
Let $\mathcal{F}$ be the class of all distribution functions on $\mathbb{R}$ and let $\mathcal{F}_i$, $i = 0,1,2...$ be a family of parametric classes of distributions on $\mathbb{R}$.

(1) Analysis in the neighborhood of a given model F_0



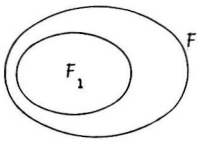
Let $\mathcal{F}_1$ be a wider class of models including F_0. Working with $\mathcal{F}_1$ we obtain more robust methods and a sensitivity analysis will be considered in this neighborhood of F_0.

(2) Mixtures of models



Working with $\mathcal{F}_0 \cup \mathcal{F}_1 \cup \mathcal{F}_2 \ldots$, allows us to obtain even more robust methods and provides a framework for analysing the sensitivity of the inference to the choice of a model.

(3) Approaches to nonparametric methods



Let $\mathcal{F}_1$ be a "large" class of models so that "almost distribution-free" methods are obtained.

It will be shown that Bayesian methods provide a natural framework for posing the problem but that two kinds of difficulties may be encountered : computational problems and problems of interpretation (choice of parametrisation; assessment of prior distribution, etc.).

2. INFERENCE ROBUSTNESS

In this section, it will be seen that the Bayesian formulation of the approaches (1) and (2) above, implies the introduction of some hyperparameters either describing the neighborhood of a given model, or tracing a family of different model. Therefore, the mathematical formulation and the difficulties are essentially the same. However, the structure of the priors will be quite different.

2.1. Neighborhood of a model

The idea is to consider the data to be generated by some member of a wider class of distribution functions F_1, including as a subset F_0 the family of distribution functions represented by the original model. The date density of F_1 will now be written $p(x|\theta,\nu)$ where θ is the parameter of interest and ν is a new parameter characterising some neighborhood of the basic model; if ν_0 is the value of ν corresponding to the basic model, then $\forall\ \theta \in \Theta$, $p(x|\theta,\nu_0)$ is a member of F_0.

Once the prior on the new parameter space $p(\theta,\nu)$ is specified, the Bayesian analysis is straightforward :

$$p(\theta,\nu|x) \propto p(x|\theta,\nu)\ p(\theta,\nu). \tag{2.1}$$

The following outputs of the analysis are of interest :

(i) $p(\theta|x)$ represents the overall (marginal) information on θ. So the inference on θ is now based on a wider class of models and this provides more "robust" procedures.

(ii) $p(\nu|x)$ allows one to analyse the plausibility of several models including the simpler model $(\nu = \nu_0)$ and it shows how the uncertainty about the model, represented by $p(\nu)$, is transformed by the data.

(iii) $p(\theta|x,\nu)$ computed for different value of ν, reveals the sensitivity of the inference about θ to small departures from $\nu = \nu_0$.

Thus, it appears that the Bayesian approach provides (at least from a theoretical point of view) the natural framework to think about robustness problems. However, there may be some practical difficulties in performing this analysis. These difficulties will be discussed after the two following examples.

Example 1 (Box-Tiao (1973)) : The family of exponential power densities

The data density (for one observation) is written as :

$$p(x|\mu,\sigma,\nu) = \frac{[\Gamma(3/2(1+\nu))]^{1/2}}{(\nu+1)\ [\Gamma(1/2(1+\nu))]^{3/2}}\ \sigma^{-1} \exp\left\{-\left[\frac{\Gamma(3/2(1+\nu))}{\Gamma(1/2(1+\nu))}\right]^{1/1+\nu} \left|\frac{x-\mu}{\sigma}\right|^{2/1+\nu}\right\} \tag{2.2}$$

where $\mu \in \mathbb{R}$, $\sigma \in \mathbb{R}^+$, $-1 < \nu \leq +1$. This may be written as

$$p(x|\mu,\sigma,\nu) \propto \sigma^{-1} \exp\left\{- a(\nu) \left|\frac{x-\mu}{\sigma}\right|^{2/1+\nu}\right\} \tag{2.3}$$

and represents a class of symmetric distributions including the normal ($\nu = 0$), the double exponential ($\nu = 1$) and the uniform (when ν tends to -1). The different posterior p.d.f. are obtained by numerical integration and Box and Tiao (1973) propose a particular form for $p(\nu)$ which facilitates these computations.

Example 2 (Relles-Rogers (1977)) : The family of student densities

$$p(x|\mu,\sigma,\nu) = \frac{\nu^{1/2}}{B\,(1/2,1/2\nu)}\,\sigma^{-1}\left[1 + \nu\left(\frac{x-\mu}{\sigma}\right)^2\right]^{-(1+\nu)/2\nu} \tag{2.4}$$

where $\mu \in \mathbb{R}$, $\sigma \in \mathbb{R}^+$, $0 \leq \nu \leq 1$. In this parametrisation, ν is the inverse of the degrees of freedom; when $\nu = 0$ we obtain the normal density and for $\nu = 1$, the Cauchy density. Relles and Rogers (1977) analyse by simulation the robustness properties of the estimator of μ obtained from $p(\mu|x)$ versus some classical robust estimators for location parameters. They pointed out the good performances of the Bayesian estimator.

Comments

The analysis presented above seems elegant and very attractive. But, in addition to the computional difficulties, there may be a lot of problems in assessing the prior p.d.f. $p(\theta,\nu)$. Usually, this prior is specified through the following decomposition :

$$p(\theta,\nu) = p(\theta|\nu)\,p(\nu) \tag{2.5}$$

and the questions are : how to assess $p(\theta|\nu)$ for different values of ν in a coherent way ? Does θ represent the same physical characteristic of the population for different value of ν ? It is reasonable to let θ and ν be prior independent ? In the latter case $p(\theta|\nu) = p(\theta)$ and the specification of $p(\theta,\nu)$ would certainly be facilitated. Thus it appears that the choice of the parametrisation seems to be crucial.

In example 1 (exponential power family) μ and σ represent the mean and standard deviation of the population for each value of ν and Box-Tiao (1973) suppose that (μ,σ) and ν are a priori independent. The following question may be raised : is it reasonable that the prior on σ is the same for $\nu = 0$ (the normal case) and for ν approaching -1 (the uniform case) ?

In example 2 (student family), the choice of the parametrisation is important. The student family could have been parametrised as follows :

$$p(x|\mu,\alpha,\nu) = \frac{1}{B\,(1/2,1/2\nu)}\,\alpha^{-1}\left[1 + \left(\frac{x-\mu}{\alpha}\right)^2\right]^{-1/2\nu - 1/2} \tag{2.6}$$

when $\alpha \in \mathbb{R}^+$ is a different choice of scale parameter.

We have the following relations between α and σ :

$$\text{Var}\,(x|\mu,\sigma,\nu) = \frac{\sigma^2}{1-2\nu} \quad \text{if} \quad \nu < 1/2; \tag{2.7}$$

$$\text{Var}\,(x|\mu,\alpha,\nu) = \frac{\nu\alpha^2}{1-2\nu} \quad \text{if} \quad \nu < 1/2. \tag{2.8}$$

The question may be raised : is it more reasonable to have a priori independence between σ and ν or between α and ν ? (Relles-Rogers (1977) have chosen $p(\sigma,\nu) = p(\sigma)$). One could have chosen $\gamma = \sigma^2/1-2\nu$ as parameter (with $\nu < 1/2$). In this case μ and γ would have the same interpretation for all ν but the problems of independence are still present. Note that Box and Tiao (1973) have pointed out that in the location scale family, a scale parameter is arbitrarily defined up to a multiplicative constant. If $\alpha = f(\nu)\,\sigma$, the prior independence between σ and ν is lost for α and ν.

More generally, it seems reasonable that prior opinion about θ must influence one's opinion about the possible models. This appears when writing $p(\nu|x)$ even if $p(\theta,\nu) = p(\theta)\,p(\nu)$, since

$$p(\nu|x) \propto p(\nu) \int_{\Theta} p(x|\theta,\nu)\; p(\theta)\; d\theta. \tag{2.9}$$

No general answers can be given to this kind of question, but the comments above show that the elegance of the theoretical development should not hide the practical problems of interpretation.

2.2. Mixtures of models

Another way of enlarging the class of models considered is to extend the ideas of the preceding section in the following way; let the data density be written as follows

$$p(x|\theta_\lambda,\lambda) \tag{2.10}$$

where $\theta_\lambda \in \Theta_\lambda$, $\lambda \in \Lambda$, and where, for each $\lambda \in \Lambda$ (Λ is typically discrete), $p(x|\theta_\lambda,\lambda)$ is a particular parametric density with parameter space Θ_λ. Here again, in addition to the prior densities $p(\theta_\lambda|\lambda)$ a prior probability $p(\lambda)$ for each model must be specified. As before the Bayesian analysis is elegant and straightforward :

$$p(x|\lambda) = \int_{\Theta_\lambda} p(x|\theta_\lambda,\lambda)\; p(\theta_\lambda|\lambda)\; d\theta_\lambda. \tag{2.11}$$

We can obtain

$$p(\theta_\lambda|\lambda,x) = \frac{p(\theta_\lambda|\lambda)\; p(x|\theta_\lambda,\lambda)}{p(x|\lambda)}\,. \tag{2.12}$$

We also have the posterior probabilities for each model :

$$p(\lambda|x) = \frac{p(\lambda)\; p(x|\lambda)}{\sum\limits_{\lambda\in\Lambda} p(\lambda)\; p(x|\lambda)}\,. \tag{2.13}$$