Newsletter of the European Working Group
"Multicriteria Aid for Decisions"

Bulletin du Groupe de Travail Européen
"Aide Multicritère à la Décision"

Groupe de Travail Européen "Aide Multicritère à la Décision"
Série 3, nº17, printemps 2008.

European Working Group "Multiple Criteria Decision Aiding"
Series 3, nº 17, Spring 2008.

Forum

(Robustness Analysis)

Bayesian Robustness

Fabrizio Ruggeri
CNR IMATI, Milano, Italy
(National Research Council, Institute of Applied Mathematics and Information Technology)

The interest for the Bayesian approach is growing, not only among mathematical statisticians but also among scientists and practitioners from different fields. One of the reason for the interest resides in the possibility of performing inferences or making forecasts not only based on data from statistical experiments but also on expert's knowledge. The formal combination of the two sources of information is via Bayes Theorem. I am not going to illustrate the Bayesian approach thoroughly; I refer the interested reader to the many books on specialised or general aspects of Bayesian statistics. I just mention the books by Bernardo and Smith (1994), Robert (2001) and Congdon (2006) as representative of different approaches. Application of Bayesian methods has been favoured by the burgeoning development of simulation techniques, mostly Markov chain Monte Carlo (MCMC) ones, which has made possible sampling from posterior distributions even when their mathematical expressions are not known in closed form. A recent review on simulation techniques in Bayesian statistics is provided by Gamerman and Lopes (2006).

Despite of the growing interest for Bayesian methods and, actually, emphasised by it, statisticians and scientists should make a wise use of them and be wary of the critical aspects of the approach. Here I just want to mention a relevant one: robustness. The typical Bayesian approach combines prior distributions, models and loss functions to estimate parameters, test hypotheses and forecast future observations. Given a model described by a random variable X, with density $f(x \mid \theta)$, the expert provides information which is translated into a prior distribution $\pi(\theta)$, on the parameter θ, which, combined with a sample $\mathbf{X}=(X_1, \ldots, X_n)$ from X, leads to a posterior distribution $\pi(\theta \mid \mathbf{X})$. Inferences and forecasts are based on the posterior distribution; in particular the parameter θ is estimated by specifying a loss function $L(\theta, a)$ and choosing the value of a minimising the expected posterior loss. Critics of the Bayesian approach are pointing out the arbitrariness of the choice of the prior distribution on the parameter and their concern can be extended to loss functions as well, whereas the choice of a model is a critical aspect shared by all statistical approaches. I will not discuss much about model selection and sensitivity to the choice of the model, preferring to concentrate on the two typical aspects of the Bayesian approach: prior and loss function (the former in particular). I just mention that dependence on a parametric model can be weakened by considering a Bayesian nonparametric approach, i.e. when the model does not belong to a (parametric) class but it is chosen by a probability measure on the space of all possible models. More details can be found in Ghosh and Ramamoorthi (2003).

Prior to the choice of a distribution on the parameter θ, elicitation of expert's opinion has to be performed. Different methods have been proposed, ranging from direct specification of prior distributions to lotteries and qualitative judgements, transformed into quantitative values by, e.g., controversial methods like the Analytic Hierarchy Process. Currently Dey and Ruggeri (2008) are developing a method to assess prior distributions based upon opinions on quantiles of the distribution of X, deemed as more natural than the specification of quantiles on the distribution of the parameter, especially in problems where generalised extreme value models are considered. Either directly or indirectly, expert's opinions lead to specification of some features of the prior distribution, like moments (in general, mean and variance) or quantiles. Based upon these features, the statistician chooses a suitable functional form for the prior distribution and its hyperparameters to better fit the expert's assessed values. Graphical tools, like the ones presented in Dey and Liu (2007), are helpful in showing the expert the shape of the prior corresponding to his/her assessment and induce him/her to refine it.

The choice of the prior used to be driven by mathematical considerations, like the need to get posterior distributions of known functional forms; conjugate priors are such a typical example since prior and posterior distributions have the same form (e.g. Gamma distributions when the model is either exponential or Poisson). Nowadays, powerful computers and, above all, MCMC techniques allow for almost any prior choice.

It is clear (even to supporters of the Bayesian approach) the high grade of arbitrariness affecting both elicitation and prior choice. Multiple experts' judgements are sometimes conflicting and they can hardly be combined into a unique prior. Even in the case of a unique expert, he/she might provide more quantiles than needed to determine the hyperparameters of a prior with selected functional form: some quantiles are used to find the hyperparameters and the others are used to check consistency of the assessed quantiles and the prior. At the end, a prior is chosen to fit the quantiles as well as possible. In general, identification of a unique prior corresponding exactly to expert's opinions is impossible: for a real valued parameter, the expert should be able to assign probability to any Borel set! Therefore, any prior distribution chosen by the statistician does not reflect exactly the expert's opinion. The critical implication of such arbitrariness is the influence of inaccurately specified priors on the quantities of interest, like posterior set probabilities and means. For this reason, Bayesian methods are not accepted as standard practice by many regulatory agencies in charge of, e.g., stating the efficacy of some drugs and treatments. See an interesting discussion on the topic in the *Journal of the National Cancer Institute*, Vol. 98, No. 21, November 1, 2006.

Bayesian robustness has been developed mainly to cope with this arbitrariness. The key idea behind it is the need to base inferences only on the actual assessment by the experts, specifying a class of priors compatible with their opinions and studying the influence of changes in the prior on the values of the quantity of interest. Good is among the first raising the issue of robustness, and his thoughts are discussed in Berger (1990). Robustness is considered in one of the very first Bayesian book, by Box and Tiao (1973) but it is meant about the choice of robust priors which are less affected by outliers; the choice of a t-distribution instead of a Gaussian one is the typical example. This approach is rather close in spirit, but not in mathematical tools, to the classical one which was very well described in the pioneer book by Huber (1981). I will mention later some tools, developed in the classical robustness, which have been widely used by Bayesians as well. The first relevant contributions, in the direction I am going to illustrate, are due to Kadane and Chuang (1978) and Berger (1984, 1985). The first paper illustrates stability of decision problems, specified by the triple (π, l, L) with prior π, likelihood l and loss L, under convergence of priors and losses to π and L, respectively. Philosophical aspects are illustrated in the paper by Berger (1984), whereas the first extensive discussion of the robust Bayesian approach in a textbook is due to Berger (1985). It is evident, even from the very first works on Bayesian robustness, as the interest rests mostly on prior distributions, rather than on model and loss function. The reasons for such selection are both practical (computations with classes of priors are easier than the ones with classes of models and losses) and, above all, deeper, to the very nature of the Bayesian approach, as mentioned earlier.

Berger (1990) traces back to Good's work the first interest for Bayesian robustness and discusses desiderata about classes of priors; his paper has been the reference paper for most of the people who started working in this field. The first half of the 90's saw a plethora of publications in Bayesian robustness and I was deeply involved in the two International Workshops on Bayesian Robustness, held in Italy in 1992 and 1995, whose selected papers were published in *Journal of Statistical Planning and Inference* (vol. 40, 2 & 3, 1994) and Berger et al. (1996). These proceedings and, even more, the paper by Berger (1994) discussed by the leading experts in Bayesian robustness, describe the state-of-the-art of early 90's. After the development of many methods, researchers moved to other areas of interest, especially the very innovative MCMC methods which would have changed dramatically the impact of Bayesian methods in statistics and in science. The book edited by Rios Insua and Ruggeri (2000) is not only the picture of what was developed in the Golden Age of Bayesian robustness but it is still the reference book in the field. In the last decade, Bayesian robustness has been recognised as important by many Bayesians and performed in a discrete number of papers. Still some work is done in the field and the forthcoming special issue of *International Journal of Approximating Reasoning*, for which I acted as Guest Editor, is very representative of the current research: applications, loss robustness and algorithms. The interested reader can find a useful guided tour through Bayesian robustness in Berger, Rios Insua and Ruggeri (2000), and a very updated (up to the year 2000) bibliography in Rios Insua and Ruggeri (2000).

After this brief history of Bayesian robustness, I illustrate its most relevant aspects, especially the mathematical ones. I will not discuss the foundational aspects of Bayesian robustness, but I just mention the papers by Berger (1984) and Rios Insua and Criado (2000). Links with imprecise probabilities can be found in Walley (1991).

An expert provides information on the prior distribution, e.g. moments, quantiles, unimodality, etc. and it is translated into a class Γ of priors sharing these features. Examples of classes are

$$\Gamma_S = \{\pi\text{: symmetric around } 0\}$$
$$\Gamma_{SU} = \{\pi\text{: symmetric and unimodal with mode at } 0\}$$
$$\Gamma_Q = \{\pi\text{: sharing some specified quantiles}\}$$
$$\Gamma_M = \{\pi\text{: with given mean and variance}\}$$

The above mentioned paper by Berger (1990) describes some features which are expected when a class of priors is specified:

- Easy elicitation and interpretation (e.g. moments, quantiles, symmetry, unimodality)
- Compatible with prior knowledge (e.g. quantile class)
- Simple computations
- Without unreasonable priors (e.g. unimodal quantile class, ruling out discrete distributions)

Once a class Γ is defined, then Bayesian robustness deals with its influence on the quantity of interest $E^* h(\theta)$, i.e. the posterior expectation of a function h of the parameter θ. The posterior mean is a typical quantity of interest since it is (see, e.g., Berger, 1995) the Bayesian optimal estimator of θ when a squared loss function $L(\theta, a)$ is considered and minimisation of posterior expected loss is chosen as the optimality criterion.

The *range*

$$\delta = \sup_\Gamma E^* h(\theta) - \inf_\Gamma E^* h(\theta)$$

is the mostly used robustness measure. When δ is *small* (according to the statistician’s and expert’s judgements), then any prior in Γ can be chosen since all of them lead to similar results. When δ is *large*, then further information is needed to get a smaller class Γ^* (as an example, further quantiles could be added). Smaller and smaller classes can be considered until either a *small* range is obtained or a *large* range cannot be further reduced. In the latter case, a prior distribution in Γ can be considered and the corresponding value of $E^* h(\theta)$ reported along with the range. The choice of such prior can be driven by mathematical convenience if there is one leading to tractable computations. My favourite choice, not shared by all Bayesian statisticians, is a prior which satisfies some optimality criterion, like Γ-minimax posterior expected loss and Γ-minimax posterior regret, considered in, e.g., Betrò and Ruggeri (1992) and Rios Insua, Ruggeri and Vidakovic (1995), respectively.

Other classes of priors have been proposed in literature; they can be classified in different ways but here I would like to emphasise the distinction between those in which all priors share some features (like the one described before) and the neighbourhood classes. Many classes in the former group can modelled as generalised moments constrained classes, studied first by Betrò, Meczarski and Ruggeri (1994). The latter classes are not necessarily neighbourhood of a baseline prior distribution in a topological sense; sometimes they are just perturbations of such prior. These classes arise when a given distribution is considered a good candidate to reflect the prior knowledge and the effects of departures from it are studied in terms of changes in the quantity of interest. A typical example, proposed by Huber in the classical framework, is given by the ε-contaminations

$$\Gamma_\varepsilon = \{\pi: \pi = (1-\varepsilon)\,\pi_0 + \varepsilon\ q, q \text{ in } \Omega\},$$

where π_0 is the baseline prior and Ω a class of priors, e.g. all possible, all symmetric or all symmetric and unimodal ones. Other neighbourhood classes are based on bounds on density functions or distribution functions. A general way to introduce topological neighbourhoods of a baseline prior is illustrated in Fortini and Ruggeri (1994) who use the concentration function.

Computation of ranges in classes of priors is considered a typical example of global sensitivity analysis, since it is a measure related to a whole class. As discussed in Berger, Rios Insua and Ruggeri (2000), two other approaches play a relevant role: informal and local sensitivity. The former approach is used in many Bayesian papers when few, different priors are entertained and the inferences upon them are compared. The latter approach measures the effect of infinitesimal changes in the baseline prior, using Frechet and Gateaux derivatives as in the papers by Ruggeri and Wasserman (1993 and 1995, respectively). This approach resembles the one based on influence functions developed in classical robustness. I will not discuss these approaches any further.

Computations of ranges is a difficult task, since it involves, in principle, functional optimisation. Most of the work in the early 90’s was devoted to the search of algorithms to compute robustness measures, mostly the range. The goal was, in general, the transformation of the functional optimisation problem into a more manageable, nonlinear one. The key result behind such transformation is the equivalence, in general,

$$\sup_\Gamma E^* h(\theta) = \sup_\Delta E^* h(\theta)$$

where Δ is the set of the extremal measures in Γ, i.e. those measures such that any measure in Γ can be expressed as their mixture. As an example, Sivaganesan and Berger (1989) prove that the supremum in the ε-contamination class with Ω = {all probability measures} is achieved for a Dirac measure at some value ω of the parameter, so that the optimisation problem has to be solved just searching for the optimal ω.

Since $E^* h(\theta)$ is the ratio of two linear quantities in π, i.e. $\int h(\theta)\, l(\theta)\, \pi(\theta)\, d\theta$ and $\int l(\theta)\, \pi(\theta)\, d\theta$, with $l(\theta)$ being the likelihood function, optimisation is not a trivial task. For such reasons, Lavine (1991) proposed a linearisation technique which could make computations easier. Other approaches, proposed by O’Neill (2008) and Betrò and co-authors (see the most recent work by Betrò, 2008, for references to past works), are based, respectively, on importance sampling and linear semi-infinite programming applied to generalised moments constraints classes. The lack of multi-purpose, interactive, widely available software able to compute ranges for a large variety of classes is one of the reasons for the limited application of the plethora of methods developed in the last 15-20 years. Robustness analysis is deemed important by Bayesian statisticians but it is only sometimes performed in actual analyses, apart from some informal sensitivity check. The development of such user-friendly software is one of the major challenges ahead.

I have briefly described what is called robustness with respect to the prior, avoiding other important aspects, like loss and likelihood robustness. The former deals with classes of loss functions and the consequent range of Bayesian optimal estimators or with the search of non-dominated actions under classes of priors and/or losses, whereas the latter deals with perturbations of the likelihood. I refer to the review papers by Martin and Arias (2000) and Shyamalkumar (2000), respectively.

REFERENCES

1. Berger , J.O. (1984), The robust Bayesian viewpoint (with discussion). In Robustness of Bayesian Analysis (J. Kadane ed.), North Holland, Amsterdam.
2. Berger, J.O. (1985), Statistical Decision Theory and Bayesian Analysis, Springer-Verlag, New York.
3. Berger, J.O. (1990), Robust Bayesian analysis: sensitivity to the prior; Journal of Statistical Planning and Inference, vol. 25, pp. 303-328.
4. Berger, J.O. (1994), An overview of robust Bayesian analysis (with discussion), TEST, vol. 3, pp. 5-59.
5. Berger, J.O., Betrò, B., Moreno, E., Pericchi, L.R., Ruggeri, F., Salinetti, G. and Wasserman, L. Eds. (1996), Bayesian Robustness, Lecture Notes IMS, vol. 29, Institute of Mathematical Statistics, Hayward.
6. Berger, J.O., Rios Insua, D. and Ruggeri, F. (2000), Bayesian robustness. In Robust Bayesian Analysis (D. Rios Insua and F. Ruggeri, eds.), Springer-Verlag, New York.
7. Bernardo, J. and Smith, A.F.M. (1994), Bayesian Theory, Wiley, Chichester.
8. Betrò, B., (2008), Numerical treatment of Bayesian robustness problems, to appear in International Journal of Approximate Reasoning.
9. Betrò, B., Meczarski, M. and Ruggeri, F. (1994), Robust Bayesian analysis under generalized moments conditions, Journal of Statistical Planning and Inference, vol. 41, pp. 257-266.
10. Betrò, B. and Ruggeri, F. (1992), Conditional Γ-minimax actions under convex losses, Communications in Statistics – Theory and Methods, vol. 21, pp. 1051-1066.
11. Box, G.E.P., and Tiao, G.C. (1973), Bayesian Inference in Statistical Analysis, Wiley, New York.
12. Congdon, P. (2006), Bayesian Statistical Modelling, 2nd Edition, Wiley, Chichester.
13. Dey, D. and Liu, J. (2007), A Quantitative Study of Quantile Based Direct Prior Elicitation from Expert Opinion, Bayesian Analysis, vol. 2, pp. 137-166.
14. Dey, D. and Ruggeri, F. (2008), Model based prior elicitation, manuscript.
15. Fortini, S. and Ruggeri, F. (1994), On defining neighbourhoods of measures through the concentration function, Sankhya, Series A, vol. 56, pp. 444-457.
16. Gamerman, D. and Lopes, H. (2006), Markov Chain Monte Carlo: Stochastic Simulation for Bayesian Inference, 2nd Edition), Chapman and Hall/CRC Press, Boca Raton.
17. Ghosh, J.K. and Ramamoorthi, R.V. (2003), Bayesian Nonparametrics., Springer-Verlag, New York.
18. Huber, P.J. (1981), Robust Statistics, Wiley, New York.
19. Kadane, J. and Chuang, D.T. (1978), Stable decision problems, Annals of Statistics, vol. 6, pp. 1095-1110.
20. Lavine, M. (1991), Sensitivity in Bayesian statistics; the prior and the likelihood, Journal of the American Statistical Association, vol. 86, pp. 396-399.
21. Martin, J. and Arias, J.P. (2000), Computing efficient sets in Bayesian decision problems. In Robust Bayesian Analysis (D. Rios Insua and F. Ruggeri, eds.), Springer-Verlag, New York.
22. O’Neiil, B. (2008), Importance sampling for Bayesian sensitivity analysis, to appear in International Journal of Approximate Reasoning.
23. Rios Insua, D. and Criado, R. (2000), Topics on the foundations of robust Bayesian analysis. In Robust Bayesian Analysis (D. Rios Insua and F. Ruggeri, eds.), Springer-Verlag, New York.
24. Rios Insua, D. and Ruggeri, F. Eds.(2000), Robust Bayesian Analysis, Springer-Verlag, New York.
25. Rios Insua, D., Ruggeri, F. and Vidakovic, B. (1995), Some results on posterior regret Γ-minimax estimation, Statistics and Decision, vol. 13, pp. 315-331.
26. Robert, C. (2001), The Bayesian choice, 2nd Edition, Springer-Verlag, New York.
27. Ruggeri, F. and Wasserman, L. (1993), Infinitesimal sensitivity of posterior distributions, Canadian Journal of Statistics, vol. 21, pp. 195-203.
28. Ruggeri, F. and Wasserman, L. (1995), Density based classes of priors: infinitesimal properties and approximations, Journal of Statistical Planning and Inference, vol. 46, pp. 311-324.
29. Shyamalkumar, N.D. (2000), Likelihood robustness. In Robust Bayesian Analysis (D. Rios Insua and F. Ruggeri, eds.), Springer-Verlag, New York.
30. Sivaganesan, S. and Berger, J.O. (1989), Ranges of posterior measures for priors with unimodal contaminations, Annals of Statistics, vol. 17, pp. 868-889.
31. Walley, P. (1991), Statistical Reasoning with Imprecise Probabilities, Chapman and Hall, London.