SETS OF PRIOR PROBABILITIES AND BAYESIAN ROBUSTNESS

LUIS RAUL PERICCHI

ABSTRACT. We outline a brief review of the different approaches which use sets of probability measures as models for prior imprecise knowledge.

This is a contribution to the *Documentation Section* on the website of the *Society for Imprecise Probability Theory and Applications* (SIPTA): `http://www.sipta.org`

1. JUSTIFICATION OF SETS OF PROBABILITY MEASURES AS MODELS FOR PRIOR UNCERTAINTY

It has been shown that behaviour can be coherent without satisfying the assumption of *precision*, i.e. without the assumption that arbitrarily fine distinctions between probabilities can be made — see (Walley, 1991) [17] for such a proof and references. Coherence corresponds to subjective Bayesian behaviour, but with a less restrictive set of assumptions. In this article we focus on strategies for choosing, and ways of interpreting, classes of prior distributions, which attempt to model the unavoidable imprecision in life and nature.

There are two different broad interpretations. The first is the *sensitivity analysis* or *collection of individual priors* interpretation, which is closer to the usual Bayesian approach: classes of priors are composed of priors individually judged to be reasonable and compatible with the (partial) available information. Thus each prior is consistent with actual prior beliefs, but it is recognised that prior beliefs are imprecise. This will be called 'Bayesian robustness'. The second interpretation may be called the *collective prior* or *upper and lower probability* interpretation. Here it is the properties and features of the whole class that matter (particularly the extremes reflected in upper and lower probabilities and expectations), and the particular features of individual priors are unimportant.

The calculi of both interpretations are similar, but the assessment strategies are different. In the first approach each of the priors should be sensible subjectively, whereas in the second approach it is the whole class which matters — it is a reasonable class of priors and not a class of reasonable priors. See (Pericchi and Walley, 1991) [14] for discussion and references. Also the problems addressed by each interpretation are different: Bayesian robustness aims to establish a neighbourhood around a sensible, a priori subjective measure. On the other hand, under the second interpretation, '*near ignorance priors*' with natural invariance properties have been proposed as a robust and imprecise alternative to a standard noninformative (but precise) prior distribution. The behaviour with respect to different data is also different: neighbourhood models are a posteriori more imprecise when there is a conflict or inconsistency between the data and prior information, which cannot arise for invariant classes.

In this contribution we consider sets of measures only as models for *prior* uncertainty. This is done for the following reasons:

(1) the prior is, or at least is perceived, to be the weakest link in the logic of coherence;
(2) imprecision in the sampling model, although typically more important than imprecision in the prior, is more effectively dealt with by model selection and hypothesis testing procedures, as is usually done in statistics.

For a combination of robust Bayesian priors and model selection for models with the same dimension, see (Pericchi and Pérez, 1994) [13]. For models with different complexity, priors will always have a pronounced effect. Priors across different models should be reasonably 'tied together'. Priors should not be completely unrelated for different models. For discussion and developments see (Berger and Pericchi, 1996) [3] and (Sansó, Pericchi and Moreno, 1996) [16].

2. Bayesian robustness with respect to the prior

The following factors in choosing a class of priors were identified by Berger (1994) [2]. The class should be:

- easy to elicit and interpret;
- easy to handle computationally;
- big enough to reflect prior uncertainty;
- extensible to higher dimensions.

2.1. **Parametric classes of priors.**

This approach selects the class of priors Γ to be indexed by a finite dimensional parameter. Advantages of these classes are that they are simple with respect to both computation and communication. Also the assessments may be more familiar. The main disadvantage is that the class might be too narrow, not allowing enough variation in features. Still parametric classes are certainly useful, unless the deviations from the basic assumptions are completely unexpected.

Examples are classes of conjugate priors, but there are several other examples of non-conjugate parametric classes as well. Modern simulation-based algorithms are effective in computations with non-conjugate but parametric classes — see (Geweke, 1998) [7].

2.2. **Non-parametric classes of priors.**

Non-parametric classes have been more popular in recent research in the area. Partly this is due to the interesting mathematics they lead to. But the main advantage is the potentially wide range of behaviour that can be modelled with non-parametric classes. Here we mention only the main classes that have been considered — see (Berger, 1994) [2], (Wasserman, 1992) [20] and (Walley, 1991) [17] for other classes and references.

Classes around a default prior: One of the first contributions on classes of priors was the theory of precise measurement of Edwards, Lindman and Savage (1963) [6]. This in fact defines a class around a default prior. This class however is defined a posteriori and requires various assessments. Still it is a very interesting class, and some improvements to the upper and lower bounds in the original paper have been made by Moreno, Pericchi and Kadane (1998) [12].

ϵ-contamination classes: The class of priors Γ is defined by

$$\Gamma = \{(1-\epsilon)\pi_0 + \epsilon\phi : \phi \in \Phi\},$$

where π_0 is the base prior or origin, ϵ is the assessed amount of error in the base prior, and Φ is the class of contaminations considered. This class is reasonably tractable, and some variation in size is obtained by imposing constraints like unimodality and symmetry on the permissible contaminations. See (Berger, 1994) [2], (Walley, 1991) [17], (Moreno and Cano, 1991) [10] and (Moreno and Pericchi, 1993) [11] for discussion.

Density-ratio classes: These were proposed by DeRobertis and Hartigan (1981) [5]. The class is defined as:

$$\Gamma = \{\pi : L(\theta) \leq \pi(\theta) \leq U(\theta)\},$$

where the measure π is a prior which does not necessarily integrate to one. This class has many advantages, not only computational — see for example (Wasserman, 1992) [20]. Assessment of L and U can be very difficult however. Pericchi and Walley (1991) [14] have proposed choosing L to be an assessed (possibly conjugate) prior and U to be a default or 'non-informative' prior which touches $L(\theta)$ at its maximum.

Quantile classes: These have the form:

$$\Gamma = \{\pi : \alpha_i \leq \pi(\Theta_i) \leq \beta_i, i = 1, ..., m\},$$

where Θ_i are specified subsets of Θ. This is natural for elicitation but it allows very wild priors, and it is not sensible in higher dimensions unless shape constraints are introduced which make the model less tractable.

Mixture classes: These are defined as:

$$\Gamma = \{\pi : \pi(\theta) = \int \pi(\theta|\alpha) g(\alpha) \mathrm{d}\alpha, g \in G\}.$$

These classes are quite useful, particularly in high dimensions, since such a class is very smooth.

3. Collective robustness of classes

The desiderata here have points in common with the previous ones. There is one important point of difference however. Recall that these classes are intended to model prior ignorance. Thus, with respect to important aspects, the class should be as imprecise as it can be. This leads us to near ignorance classes.

3.1. Near ignorance classes.

The natural candidate for a class to represent complete ignorance is the class of all distributions. When this class leads to non-vacuous and useful conclusions, these are quite compelling and uncontroversial. An instance of this is the proof in (Berger and Sellke, 1987) [4] that p-values cannot be considered as evidence for or against a point null hypothesis. There are rather few examples of the usefulness of the class of all distributions, however. More often it turns out that the posterior probabilities obtained from this class are vacuous, i.e., their lower and upper bounds are 0 and 1. There is then a compromise to be made, and this is the compromise of *near*-ignorance. The near-ignorance class should be vacuous a priori in some respects, typically the most important ones. For instance, if we are are interested in a location parameter, the class should be translation invariant, but such that, for example, posterior expectations or posterior probabilities of intervals are not vacuous. An example in (Pericchi and Walley, 1991) [14] is the following: assume a normal location likelihood, i.e., with unknown location θ and known scale σ. A very reasonable near ignorance class for this problem turns out to be a class of double exponential priors:

$$\Delta = \{\pi : \pi(\theta) = \frac{1}{\sqrt{2}\tau} \exp\left(\frac{-\sqrt{2}}{\tau}|\theta - \mu|\right), -\infty < \mu < \infty\},$$

which is location invariant and depends upon a single assessment τ. See (Walley, 1991) [17] and (Sansó and Pericchi, 1992) [15] for other examples. Unfortunately however, the classes proposed there, although appropriate for the examples considered, still require some prior elicitation and are not fully automatic.

4. Conclusions

There are two interpretations and two broad kind of models involving sets of probability measures. The first (and more developed) is the robust Bayesian approach of neighbourhood models. The second is the collective prior approach of upper and lower probability theory, which aims to provide an imprecise alternative to objective Bayesian default analyses. This ambitious aim still needs to be developed so that it is elicitation free and can be constructed with knowledge only of the likelihood, to have more influence on the practice of statistics.

Acknowledgements

The author gratefully acknowledges a Guggenheim Fellowship and support by CONICIT-Venezuela, and thanks the IPP editors for extensive editorial assistance.

Note added by editors

This article has concentrated on the ways in which sets of probability measures can be used to model imprecise prior information in statistical problems, and it has not attempted to survey other theories which use sets of probability measures to model uncertainty. Some references to the other theories can be found in the bibliography on this site, under the heading *Sets of probability measures*. We especially want to point out the theory of Levi (1980 and later references) [9]. Other important references concerning the topics of this article are (Good, 1962) [8] and (Berger, 1984) [1] on robust Bayesian inference, and (Walley, 1996) [18] and (Walley, 1997a) [19] on near ignorance models.

References

[1] J. O. Berger. The robust Bayesian viewpoint. In J. B. Kadane, editor, *Robustness of Bayesian Analyses*, pages 63–144. North-Holland, Amsterdam, 1984. with discussion.

[2] J. O. Berger. An overview of robust Bayesian analysis. *Test*, 3:5–124, 1994. with discussion.

[3] J. O. Berger and L. R. Pericchi. The intrinsic Bayes factor for model selection and prediction. *Journal of the American Statistical Association*, 91:109–122, 1996.

[4] J. O. Berger and T. Sellke. Testing a point null hypothesis: the irreconcilability of significance levels and evidence. *Journal of the American Statistical Association*, 82:112–122, 1987. with discussion.

[5] L. DeRobertis and J. A. Hartigan. Bayesian inference using intervals of measures. *The Annals of Statistics*, 9:235–244, 1981.

[6] W. Edwards, H. Lindman, and L. J. Savage. Bayesian statistical inference for psychological research. *Psychological Review*, 70:193–242, 1963.

[7] J. Geweke. Simulation methods for model criticism and robustness analysis. In J. M. et al. Bernardo, editor, *Bayesian Statistics*, volume 6. Oxford University Press, Oxford, 1998. with discussion.

[8] I. J. Good. Subjective probability as the measure of a non-measurable set. In E. Nagel, P. Suppes, and A. Tarski, editors, *Logic, Methodology and Philosophy of Science*, pages 319–329. Stanford University Press, Stanford, 1962.

[9] I. Levi. *The Enterprise of Knowledge*. MIT Press, London, 1980.

[10] E. Moreno and J. A. Cano. Robust Bayesian analysis with ϵ-contaminations partially known. *Journal of the Royal Statistical Society, Series B*, 53:143–155, 1991.

[11] E. Moreno and L. R. Pericchi. Bayesian robustness for hierarchical ϵ-contamination models. *Journal of Statistical Planning and Inference*, 37:159–168, 1993.

[12] E. Moreno, L. R. Pericchi, and J. Kadane. A robust Bayesian look at the theory of precise measurement. In J. et al. Shantan, editor, *Decision Research from Bayesian Approaches to Normative Systems*. Kluwer Academic, 1998.

[13] L. R. Pericchi and M. E. Pérez. Posterior robustness with more than one sampling model. *Journal of Statistical Planning and Inference*, 40:279–294, 1994.

[14] L. R. Pericchi and P. Walley. Robust Bayesian credible intervals and prior ignorance. *International Statistical Review*, 58:1–23, 1991.

[15] B. Sansó and L. R. Pericchi. Near ignorance classes of log-concave priors for the location model. *Test*, 1:39–46, 1992.

[16] B. Sansó, L. R. Pericchi, and E. Moreno. On the robustness of the intrinsic Bayes factor for nested models. In *Bayesian Robustness*, volume 29 of *Lecture Notes of the Institute of Mathematical Statistics*, pages 155–174. 1996.

[17] P. Walley. *Statistical Reasoning with Imprecise Probabilities*. Chapman and Hall, London, 1991.

[18] P. Walley. Inferences from multinomial data: learning about a bag of marbles. *Journal of the Royal Statistical Society, Series B*, 58:3–57, 1996. with discussion.

[19] P. Walley. A bounded derivative model for prior ignorance about a real-valued parameter. *Scandinavian Journal of Statistics*, 24:463–483, 1997.

[20] L. A. Wasserman. Recent methodological advances in robust Bayesian inference. In J. M. Bernardo, J. O. Berger, A. P. Dawid, and A. F. M. Smith, editors, *Bayesian Statistics*, volume 4, pages 483–502. Oxford University Press, Oxford, 1992. with discussion.