Journal of Statistical Planning and Inference 141 (2011) 382–391

Contents lists available at ScienceDirect

Journal of Statistical Planning and Inference

journal homepage: www.elsevier.com/locate/jspi

Considerations on jury size and composition using lower probabilities

Frank P.A. Coolen [a,*], Brett Houlding [b], Steven G. Parkinson [c]

[a] Department of Mathematical Sciences, Durham University, UK
[b] Department of Statistics, Trinity College Dublin, Ireland
[c] Department of Engineering Mathematics, University of Bristol, UK

ARTICLE INFO

Article history:
Received 25 August 2009
Received in revised form
1 June 2010
Accepted 10 June 2010
Available online 12 June 2010

Keywords:
Jury size
Lower and upper probability
Representative decision making
Stratified sampling
Subpopulations

ABSTRACT

The use of lower probabilities is considered for inferences in basic jury scenarios to study aspects of the size of juries and their composition if society consists of subpopulations. The use of lower probability seems natural in law, as it leads to robust inference in the sense of providing a defendant with the benefit of the doubt. The method presented in this paper focusses on how representative a jury is for the whole population, using a novel concept of a second 'imaginary' jury together with exchangeability assumptions. It has the advantage that there is an explicit absence of any assumption with regard to guilt of a defendant. Although the concept of a jury in law is central in the presentation, the novel approach and the conclusions of this paper hold for representative decision making processes in many fields, and it also provides a new perspective to stratified sampling.

1. Introduction

In law, the use of juries is often regarded as a natural manner for reaching a verdict, mostly used when a defendant is charged with a serious crime. In such situations, there is typically uncertainty about the guilt of the defendant, and most civilized societies only wish to convict the defendant if there is considered to be very strong evidence that the defendant committed the crime: in case there is remaining doubt, the defendant should normally be given the benefit of the doubt, and should not be convicted. Due to the presence of uncertainty, it is natural that probabilistic and statistical methods have been used to analyze several theoretical aspects of juries (e.g. Friedman, 1972), and of uncertainty in law more generally (e.g. Gastwirth, 2000). During a trial, an enormous amount of information is typically presented to a jury. Such information may consist of many facts brought alight, with different emphasis on their relevance and circumstances under which these facts did occur (or not) or might have occurred (or not), and the manner in which this all is presented can be confusing to members of the jury. Clearly, this makes it difficult to translate all such information into suitable data for a statistical approach based on a full probabilistic model, and the one-off nature of specific court cases appears to prevent a classical frequentist statistical approach to support jurors in reaching a verdict. From a Bayesian perspective, it would be extremely difficult to provide a detailed model at the prior stage, as one would have to foresee all possible information that might appear in a court case, in the right order (as e.g. the defence will often adapt its strategy to counter arguments presented by the prosecutor), and based on detailed expert judgements (as, effectively, only one realization of the whole process is actually observed, so any prior information is likely to remain influential).

* Corresponding author.
E-mail addresses: frank.coolen@durham.ac.uk (F.P.A. Coolen), houldinb@tcd.ie (B. Houlding), steven.parkinson@bristol.ac.uk (S.G. Parkinson).

0378-3758/$ - see front matter
doi:10.1016/j.jspi.2010.06.014

Some aspects of 'uncertainty in law' have been discussed frequently, e.g. the so-called 'prosecutor's fallacy', which is a mistake due to confusion of conditional probabilities (Aitken, 2000). If one would wish to use Bayesian statistical reasoning to decide on a defendant's guilt, one would also require prior probabilities on his guilt. It would not only be very difficult to assess such prior probabilities meaningfully, but any explicit quantification of a juror's prior beliefs that the defendant is guilty would be considered to be highly inappropriate. Also, jurors are typically not trained in law, statistics, or probability; so such an approach would be deemed to fail even if suggested. It is, therefore, very difficult to even consider a suitable general way in which statistics could assist jurors with their possibly very difficult task, namely that of deducing whether or not the defendant is guilty on the basis of all evidence presented.

In this paper, we are certainly not attempting the impossible. However, we emphasize the complexity of the use of statistical methods to support jurors on deciding their verdict, as any such use of statistics is explicitly absent in the approach presented in this paper. We do not propose a method for quantifying a 'level of certainty about guilt', and we do not require any prior thoughts about the defendant's guilt. We focus our attention on juries, and we study size of juries from a novel perspective, from which we also consider composition of juries if a population consists of recognized subpopulations. The main novelty in our approach is that no assumptions are made about the defendant's guilt, and also no attempt is made to model the complex stream of information jurors have to consider during a trial. By considering a predictive criterion, which is introduced and explained in Section 3, we can still comment meaningfully on appropriateness of jury sizes from a theoretical perspective. It is important to emphasize here that we do not take practicalities of the processes used by juries to reach an overall verdict into account (Gelfand and Solomon, 1974), we assume throughout this paper that each juror takes the evidence presented into account and reaches a decision without conferring with other jurors. Actually, our approach even allows the latter to take place, but as outcomes of such deliberations might depend on particular personality characteristics of individual jurors, it would make the appropriateness of the key exchangeability assumption underlying our approach (see Section 3 and the Appendix) less clear.

Although this paper explicitly focusses on juries in basic law scenarios, the novel approach to consider appropriateness of the jury size and composition can be applied to more general representative decision making processes. In addition, the predictive method introduced in this paper for deciding on representations of subpopulations could provide an interesting novel approach to stratified sampling. So, this paper contains several novel ideas on methods for inference presented in a specific basic law context, with the intention that their development for more general problems of decision making and statistical inference will be explored in the future.

Throughout this paper, uncertainty is quantified via lower probabilities, the corresponding upper probabilities are only mentioned briefly. A basic introduction to general aspects of lower and upper probabilities is given in Section 2, with some more detailed introduction, discussion and references in the Appendix. It is particularly attractive to use lower probabilities as, for the events considered, these effectively 'give the benefit of doubt' to the defendant. As we only consider relatively straightforward statistical models with lower and upper probabilities, we use them without many further comments in the main text of this paper, but some more explanation and references are provided in the Appendix. Section 3 presents the main results of our novel approach to inference on jury size. In Section 4 we show how this approach can be used to decide on optimal representations of 'independent' subpopulations in a jury. We end the paper with some concluding remarks in Section 5.

2. Lower and upper probabilities

Most methods in statistics and quantitative risk assessment assume that uncertainty is quantified via precise probabilities, all perfectly known or determinable, which tends to assume that very detailed information is available. However, such detailed information is often not available, due to limited time for analyses or limited knowledge about the event of interest and related circumstances. In recent decades, several alternative methods for uncertainty quantification have been proposed, we restrict attention to generalized uncertainty quantification via lower and upper probabilities, also known as 'imprecise probability' (Walley, 1991) or 'interval probability' (Weichselberger, 2000, 2001). During the last decade, the theory of lower and upper probabilities has received increased attention, resulting in interesting applications (see www.sipta.org). It is widely accepted that, by generalizing precise probability theory in a mathematically sound manner, with clear axioms and interpretations, this theory provides an attractive approach to generalized uncertainty quantification.

In classical theory, a single probability $P(A) \in [0,1]$ is used to quantify uncertainty about event A. For applications, probability requires an interpretation, the most common ones are in terms of 'relative frequencies' or 'subjective fair prices for bets'. The theory of lower and upper probability (Walley, 1991; Weichselberger, 2000, 2001) generalizes probability by using lower probability $\underline{P}(A)$ and upper probability $\overline{P}(A)$ such that $0 \le \underline{P}(A) \le \overline{P}(A) \le 1$. The classical situation, so-called 'precise probability', occurs if $\underline{P}(A) = \overline{P}(A)$, whereas $\underline{P}(A) = 0$ and $\overline{P}(A) = 1$ represents complete lack of knowledge about A. Important properties are that lower (upper) probability is superadditive (subadditive), i.e. $\underline{P}(A \cup B) \ge \underline{P}(A) + \underline{P}(B)$ ($\overline{P}(A \cup B) \le \overline{P}(A) + \overline{P}(B)$) for disjoint A and B, and the conjugacy property $\underline{P}(A) = 1 - \overline{P}(A^c)$, with A^c the complementary event to A (Walley, 1991; Weichselberger, 2000, 2001). This generalization allows indeterminacy about A to be taken into account, and lower and upper probabilities can also be interpreted in several ways. One can consider them as bounds for a precise probability, related to relative frequency of the event A, and reflecting the limited information one has about A. Alternatively, from a subjective perspective the lower