Reliability Engineering and System Safety 158 (2017) 41–49

Contents lists available at ScienceDirect

Reliability Engineering and System Safety

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

Deriving the probability of a linear opinion pooling method being superior to a set of alternatives

Donnacha Bolger*, Brett Houlding

Department of Statistics, Trinity College Dublin, Dublin 2, Ireland

ARTICLE INFO

Keywords:
Decision analysis
Expert judgment
Reliability
Risk analysis
Bayesian analysis

ABSTRACT

Linear opinion pools are a common method for combining a set of distinct opinions into a single succinct opinion, often to be used in a decision making task. In this paper we consider a method, termed the Plug-in approach, for determining the weights to be assigned in this linear pool, in a manner that can be deemed as rational in some sense, while incorporating multiple forms of learning over time into its process. The environment that we consider is one in which every source in the pool is herself a decision maker (DM), in contrast to the more common setting in which expert judgments are amalgamated for use by a single DM. We discuss a simulation study that was conducted to show the merits of our technique, and demonstrate how theoretical probabilistic arguments can be used to exactly quantify the probability of this technique being superior (in terms of a probability density metric) to a set of alternatives. Illustrations are given of simulated proportions converging to these true probabilities in a range of commonly used distributional cases.

1. Introduction

In realistic decision making scenarios DMs will commonly have access to a wide range of opinions about the true value of the uncertain aspect(s) inherent within the decision task, e.g., the value of a stock price in a month, or the amount of cars that will pass along a motorway in an hour period. In addition to these opinions she (we assume the DM to be female, a common convention) also has an opinion that she herself holds. Numerous publications discuss how an amalgamated opinion can potentially outperform several of the opinions comprising it. Bates and Granger [1] consider using a weighted sum in point estimate setting with weights inversely proportional to absolute errors from previous predictions, with Newbold and Granger [27] comparing this method to several others models using a collection of eighty financial data sets and providing strong evidence in its favour. Bunn [3] considers conjugate updating (either Beta/Binomial or Dirchlet/Multinomial) for assigning weights, with performance of this scheme appearing strong in comparision to linear and exponential methods. Several of these discussions arise within an expert judgment context, where a DM requires the knowledge of domain-specific experts to assist her in her decision making task, perhaps most notably the classical method of Cooke [6,7]. This provides a framework for assessing the respective merits of the opinions of experts, using a set of seed variables (whose true values are known to the DM but unknown to the experts) to gauge their relative reliabilities. Empirical justification has been provided for this technique by Cooke and Goossens [8] and Eggstaff et al. [13], with Clemen [4] and Flandolini et al. [14] discussing the validation method used and potential alternatives. Below we consider a more novel setting, consisting of n non-competing DMs, each of whom possesses an opinion about the inherent decision uncertainty, and is willing to combine this with those of her neighbours in the hope of increasing the accuracy of the opinion that she makes her decision with, and hence her corresponding (personal) decision quality. This environment subtly differs from the expert judgment framework but shares a similar goal, i.e., constructing a combined opinion that is as accurate as possible for use in a decision making task.

Discussion on manners by which a set of opinions can be amalgamated into a single succinct opinion are widespread, with Clemen and Winkler [5] and Genest and Zidek [18] providing details for the interested reader. A fully Bayesian framework is an attractive concept, in which a DM specifies her own prior distribution before viewing the opinions of others as data to be incorporated into a likelihood function, with the product of the two yielding her posterior distribution. Yet major issues arise with the specification of an appropriate likelihood function, and while specific forms have been suggested for specific problems there is no generalised method of supplying such a function. This likelihood function would need to entail the dependence between informaiton sources, a task which will frequently be beyond the computational scope of users. French [16] comments on the attractiveness of the concept but concedes it has vast

* Corresponding author.
E-mail addresses: bolgerdo@tcd.ie (D. Bolger), brett.houlding@tcd.ie (B. Houlding).

http://dx.doi.org/10.1016/j.ress.2016.10.008
Received 1 November 2015; Received in revised form 29 September 2016; Accepted 22 October 2016
Available online 27 October 2016
0951-8320/

implementation problems, with Clemen and Winkler [5] concurring that while the method is compelling it “is also frustratingly difficult to apply”. Morris [24–26] is another example of an attempt to implement fully Bayesian updating. However this work uncovers deep problems concerning both partial exchangeability and over-determination, unfortunately making the derived approach troublesome in practice, with Morris commenting that the likelihood function can be “extremely complicated and virtually impossible to assess in all but the simplest cases”. We comment that van Noortwijk et al. [31] is one illustration of a specific application of this Bayesian approach and how, given a set of assumptions, it may be successfully implemented. Like our own methodology it is concerned with combining and updating opinions. However it does not extend to a generalised framework, and is reliant upon the user having some knowledge about the accuracy of received opinions prior to witnessing data.

The Bayesian methodology is extremely attractive but is not suitably general for widespread application. Hence various pooling methods are commonly considered, most notably in a linear or logarithmic fashion. The various approaches have differing associated strengths and weaknesses, e.g., linear pooling obeys the marginalisation property, while logarithmic pooling obeys the external Bayesianity property, both of which are discussed in, for instance, Genest and Zidek [18]. In what follows we choose to use linear opinion pooling, for its relative simplicity, ease of interpretation, and in sticking with common practice. The obvious area of interest, upon making this choice, is in determining the appropriate weights to assign, with these weights being constrained to be non-negative and to sum to unity. Genest and McConway [17] discuss various approaches by which this can be done, depending on the setting of interest, the aims of the process, and the underlying philosophy of the individual(s) assigning weights. Our setting of interest is a dynamic one, in which each DM makes a decision, sees a return, and then repeats this process indefinitely. Hence ideally two types of learning will occur over time, with a DM modifying her own opinion about the (latent) unknown decision quantity in light of the noisy realisations of it that have been observed, and also adjusting the degree of consideration that she affords to the opinions of her neighbours, given the contrast between their predictions and the witnessed reality. This dynamicity further increases the novelness and applicability of our method. We comment that DeGroot [11] and DeGroot and Bayarri [12] discuss similar settings with weight updating, but do so in a different context to us - in their context there is one set of weights to be determined/updated at each new observation, while in our case there are n sets, one for each DM.

The problem context which we highlight and provide a solution for in what follows is a substantive one which combines the fields of dynamic reliability and expert opinion. Below we include several examples of practical risk and reliability environments of this nature, illustrating the relevance of the research undertaken and some potential areas for application. This paper can be seen as providing both a methodology (and some considerable justification for its use), as well as some theoretical results which demonstrate when it is most appropriate for use. The work that is presented within this paper is highly relevant to a wide range of realistic problems across a broad spectrum of fields that pertain to technical risk and reliability.

- A collection of stockbrokers may communicate amongst each other to predict the behaviour of a financial stock, i.e., whether its price will rise or fall. Here the risk is over their personal monetary fortunes.
- A similar problem exists in the area of weather forecasting, where it has been demonstrated that a combined (ensemble) forecast can commonly outperforms single forecasts (e.g., Gneiting and Raftery, [19]). Hence there is a need to determine the weights to assign the respective forecasts. A government may require these predictions to determine if urgent anti-flood measures should be carried out to prevent against the risk of dwellings being damaged as well as individuals being injured.
- Policy developers with a non-government organisation (NGO) which provides aid to countries in the developing world may pool their respective predictions concerning the prospective Gross Domestic Product (GDP) of these nations in order to determine the ratio by which funds should be divided. Risk in this situation can be seen as one nation being allocated too much of a particular resource while another is allocated too little (e.g., food and water, vacciniations).
- A group of computer programmers may wish to pool their beliefs about the average number of bugs occurring per one thousand lines of code, in order to aid their individual decisions on whether to release their software or to continue testing it (e.g., Wilson and McDaid, [33]). The risks are clear here: if they release too early then they may need to recall software and hence lose money, as well as suffering damage to their brand reputation.
- Nuclear power stations may wish to confer between each other as to the perceived risk of a fault occurring to assist in their decision of determining whether additional safety devices need be installed or not (e.g., Starr, [30]). Potential risk to local and national safety are evident here.
- Medical practitioners may seek the opinion of peers as to the probability of a diagnosis being correct given some symptoms witnessed, or the efficacy of a novel drug treatment (and similar problems, e.g., Cox, [9]). In cases like this the risk is defined over the future health state of patient(s) using the medicine prescribed.
- Several companies may wish to exchange their opinions on what proportion of a particular demographic (e.g., males under twenty-one) buy a particular product (e.g., computer games, laptops, jeans) to ascertain how large a quantity they should respectively produce. If the proportion is overestimated then they risk too large a quantity of the goods being produced and not sold, leading to unneccesary manufacturing cost and hence a waste of captial.

In each case outlined above there is clearly technical risk present that the users wish to avoid, with this risk being defined over either financial wealth, national safety or medical health. In the illustrated situations each individual entity supplying its probabilistic opinion (be it a single person or be it a large multinational company) will have their own personal utility function over the possible outcomes which may occur, so even if decisions are made using common beliefs different decisions may well be deemed optimal by different entities.

The outline of the remainder of this paper is as follows. Section 2 introduces the Plug-in (PI) approach, a method achieving these two forms of dynamic learning that allocates weights for a linear opinion pooling. Section 3 discusses a simulation study that has been conducted to provide some validation for this technique by comparing its performance to those of some rational alternatives in different contexts. Section 4 derives the theoretical probabilities that are estimated by the simulated proportions in the previous section, illustrating how these proportions converge asymptotically to the true probabilities of interest. We conclude with discussion regarding the relevance of this material, potential applications, and further research in Section 5.

2. The plug-in approach

Suppose we have n DMs, denoted $P_1, \ldots, P_n$, with $n \geq 2$ in non-trivial cases. Each DM possesses two things: her fully parameterised probability distribution $f_i(\theta)$ over the uncertain quantity θ, and her own subjective utility function, $u_i(r)$ which is indicative of her own personal attitude to risks and gambles as consequence of decision return r. Note we assume all DMs possess some uncertainty over the opinion that they hold, i.e., all distributions have a strictly positive variance. We denote the updated opinion of P_i, having received the opinions of her neighbours, by $\hat{f}_i(\theta)$:

$$\hat{f}_i(\theta) = \alpha_{i,1} f_1(\theta) + \cdots + \alpha_{i,i} f_i(\theta) + \cdots + \alpha_{i,n} f_n(\theta) \tag{1}$$