A NOTE ON BAYESIAN DECISION THEORY WITH IMPRECISE PRIOR PROBABILITIES

F.P.A. Coolen
Eindhoven University of Technology[1]
Department of Mathematics & Computing Science

Abstract

The Bayesian framework for statistical inference offers a possibility of taking expert opinions into account, and is therefore attractive for many practical decision problems, e.g. concerning inspection and replacement of technical systems. However, the use of a single prior distribution fails to indicate the amount of information on which subjective probabilities are based, and leads to problems when combining the opinions of several experts.

The introduction of imprecise prior probabilities solves these problems, and can lead to simpler and clearer elicitation of prior information. Recently, a semi-Bayesian theory has been suggested in which imprecise prior probabilities are updated with control of imprecision related to the amount of information. Also some problems concerning elicitation of lifetime distributions, combination of opinions, introduction of statistical models and calculation of bounds on expected loss within the theory have been analyzed and solutions proposed. In this short note, intended to serve as an eye-opener to the theory, a possible application of the concept to an age replacement problem is described.

This short note has been presented at the symposium 'Reliability: A competitive edge' of the Society of Reliability Engineers (Arnhem, October 1993). The author would like to thank Martin Newby and a referee for their suggestions to improve the presentation.

1. Current address: The Department of Mathematical Sciences, University of Durham, Science Laboratories, South Road, Durham DH1 3LE, England.

1. Introduction

The theory of imprecise probabilities (Walley, 1991) is a useful and necessary generalization of the classical theory of subjective probability (De Finetti, 1974), with lower and upper probabilities describing personal betting behaviour. A bet on event A is such that the owner receives 1 if A occurs and 0 if not (Walley (1991, section 2.2) presents a method, using lottery tickets, to overcome the well-known problems if utility is expressed in terms of money). Your *lower probability* for event A, $\underline{\mathbb{P}}(A)$, is the supremum of all prices for which you want to buy the bet, your *upper probability* $\overline{\mathbb{P}}(A)$ the infimum of all prices for which you want to sell the bet, assuming that you only want to buy or sell a bet if you expect profit. Remark that in the classical theory of subjective probability you are forced to $\underline{\mathbb{P}}(A)=\overline{\mathbb{P}}(A)$, leaving no tool to represent how certain you are. This creates the obvious problem that, for example, a single probability $\mathbb{P}(head)=1/2$ when tossing a coin is used if you have no information about the coin at all as well as if you know that the coin is perfectly symmetrical. So the forced use of a single probability destroys important information, especially if probability is used as a language to exchange knowledge about uncertain events in decision problems (French, 1986), and also creates problems in the elicitation and combination of opinions (the use of higher-order probabilities only pushes the problem ahead (Walley, 1991, section 5.10)). For a survey of literature on imprecise probabilities and a presentation of axioms we refer to Walley (1991, section 1.8 and 2.7).

A simple *degree of imprecision* about (the bet on) A is $\Delta(A)=\overline{\mathbb{P}}(A)-\underline{\mathbb{P}}(A)$. In this concept $\Delta(A)$ is assumed to depend on the amount of information about A available to you. On the basis of some simple requirements Walley (1991, section 5.3) proposes $I(A)=\Delta(A)^{-1}-1$ as a measure of the amount of information about A in the imprecise probabilities. This measure plays an important role within the model used in this paper, but although Coolen (1994, section 2.2) provides a theoretical argument in favour of this relation between imprecision and information, more research is needed. The important role of imprecision in reporting the amount of information on which betting behaviour is based makes the distinction between our theory and robust Bayesian analysis (Berger, 1990). The fact that we explicitly control imprecision, as related to information, through additional parameters in case of updating leads to a concept that can be called semi-Bayesian.

In this paper Bayesian decision theory is applied to an age-replacement problem, the lifetime of a deteriorating unit is a random variable. The probability distribution of the lifetime is assumed to depend on one parameter, and opinions of experts are entered by the use of imprecise prior probabilities, following the concept of intervals of measures (DeRobertis and Hartigan, 1981). The special form of the set of prior densities and the theory of updating these priors in the light of new information are discussed by Coolen (1994), where the information measure $I(A)$ plays a crucial role in case of updating. The model and the advantages for elicitation and combination of expert opinions are briefly discussed in the example in section 2.

The use of a set of prior probabilities does not lead to indecisiveness. The difference with the classical Bayesian decision theory is that we cannot give one value of the expected loss related to a decision, but (the sharpest) bounds on the expected loss can be calculated easily (Coolen, 1994), and the distance between these bounds depends on imprecision. To reach a decision one needs an additional criterion to compare these bounds. It is believed that the use of one precise value for expected loss related to a decision is remarkable if you do not have perfect information, which is an argument in favour of our method when decision theory is applied to practical problems.

2. An Age Replacement Problem

An age-replacement rule (e.g. Tijms, 1986, section 1.2) prescribes the replacement of a unit (system, component) upon failure or upon reaching the age $\mathcal{T}$, whichever occurs first, where $\mathcal{T}$ is a control parameter. To show the possible use of Bayesian decision theory with imprecision an example is presented, with restriction to $\mathcal{T} \in \{3,6,9,..\}$ months. For the random variable $T>0$, the lifetime of the technical unit of interest, we assume a cumulative distribution function $\mathbb{F}_T(t|\theta)$, depending on a scale parameter $\theta \geq 0$ (this restriction of the parameter space is not essential to the theory but for ease of the example). A generally accepted loss function is the expected cost per time-unit over infinite time for decision $\mathcal{T}$,

$$\mathcal{L}(\mathcal{T}|\theta) = \frac{1 + (c-1)\mathbb{F}_T(\mathcal{T}|\theta)}{\int_0^{\mathcal{T}} (1-\mathbb{F}_T(t|\theta))\,dt}.$$

The cost of preventive replacement is assumed without loss of generality to be 1 and $c>1$ is the relative cost of corrective replacement. The assumption that c is precisely known may not be realistic, but generalization of the theory to $c \in [c_\ell, c_u]$ is quite easy because the comparison of the decisions $\mathcal{T}$ is based on the bounds of the expected value of $\mathcal{L}(\mathcal{T}\,|\,\theta)$ over the possible distributions for θ, and in case of $c_\ell \neq c_u$ the lower bound is accepted for costs c_ℓ and the upper bound for costs c_u.

Under the above assumptions the remaining problem is that θ is unknown, and we assume that the only information available is expert opinion about the lifetime T. Bayesian decision theory (Lindley, 1973; French, 1986) offers a means to reach a decision if a distribution for θ is given, as well as a method for updating this distribution in light of new information. However, as θ is unobservable we propose to elicit the opinions of the experts by asking questions about T, and we give possible results of such an elicitation process, together with some methods to combine these opinions and to translate the information about T into sets of prior distributions for θ. It is important for practical application of this (as any) concept of decision making that elicitation is studied in real-life cases by groups of researchers from several disciplines. So far we know of only one case-study of a decision problem using imprecision (Walley and Campello de Souza, 1990) where the concept is used as sensitivity analysis rather than to relate imprecision to the amount of information available. This last interpretation of imprecision is especially interesting if new information becomes available, which was not discussed in that case-study.

If a single prior distribution $\pi(\theta)$ is assumed, the optimum decision is that $\mathcal{T}$ which minimizes the expected loss

$$EL_\pi(\mathcal{T}) = \int_0^\infty \mathcal{L}(\mathcal{T}\,|\,\theta)\pi(\theta)d\theta.$$

Here, instead of one prior a set Π of probability densities is assumed. For each possible decision $\mathcal{T}$ let

$$\mathcal{EL}_\Pi(\mathcal{T}) = \{EL_\pi(\mathcal{T})\,|\,\pi \in \Pi\}$$

be the set of possible values of expected loss. To choose an optimal decision we compare these sets using only the lower expectation

$$\underline{\mathcal{EL}}_\Pi(\mathcal{T}) = \inf \mathcal{EL}_\Pi(\mathcal{T})$$

and the upper expectation

$$\overline{\mathcal{EL}}_\Pi(\mathcal{T}) = \sup \mathcal{EL}_\Pi(\mathcal{T}).$$

Using the intervals of measures method, we restrict Π to the form:

$$\Pi = \{ \pi \mid \pi(\theta) = q(\theta)/C_q, \quad 0 \le \ell(\theta) \le q(\theta) \le u(\theta) \text{ for all } \theta, \quad C_q = \int_0^\infty q(\theta)d\theta \},$$

where lower prior density ℓ and upper prior density u are given. Remark that in our terminology a density does not necessarily integrate to 1, whereas a probability density does. To avoid some marginal problems we restrict the discussion to ℓ and u such that

$$0 < \int_0^\infty \ell(\theta)d\theta \le \int_0^\infty u(\theta)d\theta < \infty.$$

Further, for ease of calculation when updating, ℓ is assumed to be a member of a conjugate family, and u-ℓ is also assumed to be proportional to a member of such a family. The numerical calculation of the corresponding $\underline{\mathcal{EL}}_\Pi(\mathcal{T})$ and $\overline{\mathcal{EL}}_\Pi(\mathcal{T})$ is easy (Coolen, 1994, section 3.3). Interpretation of the set Π is possible only through the related imprecise predictive cumulative distribution functions (cdf's) as used in the following example.

Example

For this example gamma distributions are used, they are mathematically attractive and describe the randomness of lifetimes of deteriorating units reasonably well. We further assume that the shape parameter of the gamma distribution of T is equal to 3, leaving a one-dimensional non-negative scale parameter θ in the model;

$$\mathbb{F}_T(t \mid \theta) = \Gamma_{\theta t}(3)/2,$$

and the probability density function is

$$f_T(t \mid \theta) = \theta^3 t^2 \exp(-\theta t)/2.$$

Completing the model in the context of Bayesian theory with imprecise prior probabilities, we assume lower and upper prior densities

$$\ell(\theta) = \tau_\ell^{10} \theta^9 \exp(-\tau_\ell \theta)/\Gamma(10)$$

and

$$u(\theta) = \ell(\theta) + c_0 a(\theta),$$

with $c_0 \ge 0$, and

$$a(\theta) = \tau_a^{10} \theta^9 \exp(-\tau_a \theta)/\Gamma(10).$$

These gamma distributions are conjugate priors for the gamma scale parameter θ, leaving hyperparameters τ_ℓ, τ_a and c_0 in the model to be chosen to make it fit well with expert opinions. To this end we compare lower and upper cdf's resulting from elicitation with the ones resulting from the model. The imprecise cdf's for T resulting from the model are (Walley, 1991, section 4.6):

$$\underline{\mathbb{F}}_T(t)=\frac{\int_0^t \ell_T(x)\,dx}{\int_0^t \ell_T(x)dx + \int_t^\infty u_T(x)dx}$$

and

$$\overline{\mathbb{F}}_T(t)=\frac{\int_0^t u_T(x)\,dx}{\int_0^t u_T(x)dx + \int_t^\infty \ell_T(x)dx},$$

where ℓ_T and u_T are Bayesian predictive densities based on the priors and therefore depending on the hyperparameters:

$$\ell_T(t)=\int_0^t f_T(t\,|\,\theta)\ell(\theta)d\theta$$

and

$$u_T(t)=\ell_T(t)+c_0 a_T(t)$$

with

$$a_T(t)=\int_0^t f_T(t\,|\,\theta)a(\theta)d\theta.$$

It is also possible to drop the assumption that the shape parameters of these gamma priors are known, but this leads to more hyperparameters and more calculations when fitting the model to subjective data. In practical applications it would be sensible to perform sensitivity analyses with regard to these assumptions.

For this example, we assume that there is one Decision Maker who wants to know the opinions of three experts. Walley and Campello de Souza (1990) suggest to use imprecise cdf's for T in the elicitation process, and we continue with the example assuming that we obtained the following results ($\underline{\mathbb{F}}_T(0)=\overline{\mathbb{F}}_T(0)=0$ and $\underline{\mathbb{F}}_T(\infty)=\overline{\mathbb{F}}_T(\infty)=1$):

	t:	6	12	18	24
Expert A:	$\underline{\mathbb{F}}_T^A(t)$:	.04	.22	.46	.66
	$\overline{\mathbb{F}}_T^A(t)$:	.15	.42	.68	.83
Expert B:	$\underline{\mathbb{F}}_T^B(t)$:	.05	.20	.41	.60
	$\overline{\mathbb{F}}_T^B(t)$:	.17	.42	.63	.78
Expert C:	$\underline{\mathbb{F}}_T^C(t)$:	.11	.32	.55	.80
	$\overline{\mathbb{F}}_T^C(t)$:	.47	.76	.89	.96

The correct interpretation of these numbers is in terms of the above betting theory for events of the type $T \leq t$, but for ease of thought one may also think that, based on the amount of information available to expert *A* at this moment, he thinks that $0.04 \leq \mathbb{P}(T \leq 6) \leq 0.15$ and does not want to make any further distinction. Experts *A* and *B* have similar ideas about the lifetime of the unit, whereas expert *C* is much more pessimistic and also less sure, which can be seen from the imprecision in the above numbers. If these numbers are interpreted using betting behaviour, both experts *B* and *C* would be pleased by a bet on the event $T \leq 24$ for price 0.79 (where *B* sells the bet to *C*).

To fit the model to subjective data of this kind (per expert)suitable values for the hyperparameters are to be determined. To reduce the amount of numerical calculations c_0 is set equal to $2\Delta_{max}/(1-\Delta_{max})$, with Δ_{max} the maximum of the imprecision for an event $T \leq t$ according to the subjective data (Coolen, 1994). Thereafter values for τ_ℓ and τ_a are determined such that the imprecise cdf's fit well to the expert's cdf's in the points where these are given. Here the distance between the lower cdf's (and for the upper cdf's) is defined as the expected squared distance of the discretized cdf's (over the intervals used in elicitation), where the expectation is with regard to the subjective lower distribution. Then τ_ℓ and τ_a were determined by minimization of the sum of these expected squared distances of the upper and of the lower cdf's. The hyperparameters for the models that fit to the above expert's cdf's are:

	c_0	τ_ℓ	τ_a
Expert A:	0.56	57.6	52.2
Expert B:	0.59	66.9	43.2
Expert C:	1.57	48.7	27.2

To calculate the bounds on the expected loss according to the above theory, the cost of corrective replacement is set to c=10. The results are (for decisions $\mathcal{T}$, with $\mathcal{T}$=300 months is effectively no replacement):

	$\mathcal{T}$:	3	6	9	12	15	18	21	24	27	30	36	300
Expert A:	$\underline{\mathcal{EL}}_\Pi$:	.34	.18	.13	.11	.10	.10	.10	.10	.10	.10	.11	.16
	$\overline{\mathcal{EL}}_\Pi$:	.34	.18	.14	.12	.12	.12	.13	.13	.13	.14	.15	.20
Expert B:	$\underline{\mathcal{EL}}_\Pi$:	.34	.17	.13	.10	.09	.09	.09	.09	.09	.09	.10	.14
	$\overline{\mathcal{EL}}_\Pi$:	.34	.18	.14	.12	.12	.12	.12	.13	.13	.14	.14	.19
Expert C:	$\underline{\mathcal{EL}}_\Pi$:	.34	.18	.14	.12	.11	.11	.12	.12	.13	.13	.14	.19
	$\overline{\mathcal{EL}}_\Pi$:	.36	.23	.22	.23	.24	.26	.27	.28	.29	.30	.31	.35

From such a scheme it is obvious that an additional criterion is needed to reach a final decision, for example, one could choose a decision per expert that minimizes the maximum expected loss, where the problem arises that these will often differ per expert. Nevertheless, from the above scheme it seems that there are some good arguments in favour of decisions $\mathcal{T}$=12 or $\mathcal{T}$=15, but the final choice is for the Decision Maker.

Another approach is to combine the information from the experts first, and then perform one analysis based on these combined opinions. Some methods are (weighted) averaging the values per expert (there is no theory of weights for combining imprecise probabilities) and two combination rules mentioned by Walley (1991, section 4.3), that result directly from the betting interpretation of the imprecise probabilities. The first one defines the new lower probability as the minimum of the lower probabilities per expert, and the new upper probability as the maximum of the upper probabilities per expert. This lower probability can be interpreted as the supremum of the prices for which all members of the group are willing to buy the bet. The second one defines the new lower probability as the maximum of the lower probabilities per expert, and the new upper probability as the minimum of the upper probabilities per expert, with interpretation that the group wants to buy or sell a bet if at least one member of the group wants this. Note that this second method of combination can lead to incoherent group betting behaviour since there may be a price at which the group would be willing to buy as well as to sell the same bet (in our example this would be caused by the disagreement between experts B and C on the event $T \leq 24$). The fact that this method actually indicates such disagreement between experts is useful in practice, and is not provided by the classical Bayesian theory (where experts always disagree, except if they assess exactly the same precise values).

An important consideration for decision theory is the possibility of incorporating new information. The updating methodology of the classical Bayesian framework is adopted here, but as this is not suitable for updating imprecision something more is needed, where the information measure plays an important role. In case of additional data consisting of n observed independent failures of the technical unit, with failure times t_i (i=1,..,n) and total time on test $tt=\sum_{i=1}^{n} t_i$, Coolen (1994) proposes updating by replacement of the hyperparameters of the densities ℓ and a according to the classical Bayesian theory (here the choice of conjugate densities leads to simple calculations) together with replacing c_0 by

$$c_n = \frac{c_0}{1+n/\xi},$$

where the additional parameter ξ is to be chosen by the DM, and can be interpreted as the amount of additional data that provides an equal amount of information as the prior (subjective) information does. Both parameters c_0 and ξ relate to imprecision, and so to the amount of information, and can be quite easily assessed, and since c_0 is related to the prior imprecision, and ξ to the value of new data compared to the subjective data, we conjecture that imprecision cannot be correctly taken into account with less than two additional parameters.

To emphasize the possibilities of this updating theory we end this example with another set of hypothetical subjective data with much imprecision, and analyse the updates in light of new information for two cases:

(I) n=10, tt=150;

(II) n=20, tt=180.

If the subjective data are

	t:	6	12	18	24
Expert D:	$\underline{\mathbb{F}}_T^D(t)$:	.03	.17	.34	.54
	$\overline{\mathbb{F}}_T^D(t)$:	.51	.76	.89	.96

the hyperparameters to fit the model to these data are

$$c_0 = 2.92, \quad \tau_\ell = 69.8, \quad \tau_a = 31.5.$$

We further assume that the weight of new information compared to the prior information is indicated by $\xi=5$. A table of the bounds on expected loss, for the prior situation and the two posterior situations is then:

		$\mathcal{T}$: 3	6	9	12	15	18	21	24	27	30	36	300
Expert D:	$\underline{\mathcal{EL}}_\Pi$:	.34	.17	.12	.10	.09	.08	.08	.08	.08	.08	.09	.13
	$\overline{\mathcal{EL}}_\Pi$:	.36	.23	.22	.22	.23	.25	.26	.27	.28	.29	.30	.33
(I):	$\underline{\mathcal{EL}}_\Pi$:	.34	.18	.13	.11	.11	.11	.11	.11	.12	.12	.13	.19
	$\overline{\mathcal{EL}}_\Pi$:	.34	.18	.14	.13	.13	.13	.14	.14	.15	.16	.17	.23
(II):	$\underline{\mathcal{EL}}_\Pi$:	.34	.20	.17	.16	.17	.19	.20	.21	.22	.23	.25	.30
	$\overline{\mathcal{EL}}_\Pi$:	.35	.21	.18	.19	.20	.21	.23	.24	.26	.27	.29	.33

It is seen that the data reduce uncertainty strongly. Again, a Decision Maker can use such a scheme, that explicitly reports lack of perfect information, to arrive at a final decision.

3. Comments

In this paper the possible use of a semi-Bayesian decision theory with imprecise probabilities is briefly shown, and some attractive features of the framework are presented. The paper is only meant to serve as an eye-opener to new ideas about the role of imprecision related to the amount of information available, and the way such a relation can be exploited in a Bayes-like theory to update imprecise prior probabilities. For many technical details we refer to recent literature. It is obvious that practical application is necessary for insight into the real value of the suggested theory as well as to solve some important open questions, e.g. about elicitation and combination of imprecise probabilities.

References

Berger, J.O. (1990) Robust Bayesian analysis: Sensitivity to the prior, *Journal of Statistical Planning and Inference* 25, 303-328.

Coolen, F.P.A. (1994) *Statistical Modeling of Expert Opinions using Imprecise Probabilities* (PhD Thesis, Eindhoven University of Technology; available upon request).

DeRobertis, L. and Hartigan, J.A. (1981) Bayesian inference using intervals of measures, *Annals of Statistics* 9, 235-244.

De Finetti, B (1974) *Theory of probability (vol. 1 and 2)* (Wiley, New York).

French, S. (1986) *Decision theory: An introduction to the mathematics of rationality* (Ellis Horwood, Chichester).

Lindley, D.V. (1973) *Making decisions* (Wiley, London).

Tijms, H.C. (1986) *Stochastic modelling and analysis: A computational approach* (Wiley, New York).

Walley, P. (1991) *Statistical reasoning with imprecise probabilities* (Chapman and Hall, London).

Walley, P. and Campello de Souza, F.M. (1990) Uncertainty and indeterminacy in assessing the economic viability of energy options: A case study of solar heating systems in Brazil, *Energy Systems and Policy* 14, 281-304.

Ontvangen: 28-2-1994
Geaccepteerd: 18-5-1994