Proc. 6th Int'l Conf. on Tools with Artificial Intelligence, IEEE Computer Society Press 1994 (pp. 668-671)

Imprecise Probability and Expert Forecasting

Carl G. Wagner

Department of Mathematics
University of Tennessee
Knoxville, TN 37996-1300

Abstract

Evidence is often insufficient to support the assessment of precise probabilities. Fortunately, shifting to vaguer measures of uncertainty, such as upper and lower probabilities, does not deprive one of the key analytical tools of classical probability. Two approaches to the calculation of upper and lower expected values are described and contrasted in the case of forecasting production costs of an electric utility. Conditionalization of imprecise probabilities is also discussed.

1 Introduction

Among the techniques used to treat uncertainty in integrated resource planning *probabilistic analysis* requires the most extensive quantification. In such an analysis, "Probabilities are assigned to different values of key uncertain variables, and outcomes are identified that are associated with the different values of the key factors in combination. Results include the expected value and cumulative probability distribution for key outcomes, such as electricity price and revenue requirements." [3].

A survey of the integrated resource plans of a number of electric utilities reveals that fully developed probabilistic analyses of uncertainty are rarely pursued. To make this observation is in no way to discredit the planning staffs of these companies. Their abstention from probabilistic analyses is a principled response to the unrealistically precise quantification of uncertainty demanded in such analyses. On the other hand, planners typically possess information that would support the assessment of upper and lower bounds, or other such constraints, on the probabilities of values of key variables.

Remarkably, shifting to vaguer assesssments of uncertainty does not deprive one of important analytical tools such as expected value and conditionalization. Useful, rigorously justified versions of both of these tools may be developed within the *theory of imprecise probabilities*, an extended account of which appears in the research report [13]. The present paper outlines some key ideas from that report.

2 Upper and Lower Probability

If Ω is a set of possible states of the world, uncertainty about which state $\omega \in \Omega$ is the true state is often modeled by a *probability measure* P defined on some class of subsets (called *events*) of Ω. The number $P(A)$ assigned to the event A represents the probability that the true state of affairs belongs to the set A. Axiomatic accounts of probability theory always postulate that $0 \leq P(A) \leq 1$ for all events A, with $P(\Omega) = 1$. In addition, *additivity* of $P(A \cap B = \emptyset \Rightarrow P(A \cup B) = P(A) + P(B))$ is always postulated, and *countable additivity* ($P(A_1 \cup A_2 \cup \cdots) = P(A_1) + P(A_2) + \cdots$, for every infinite sequence $A_1, A_2, \ldots$ of pairwise disjoint events) is often postulated (always, among mathematicians). As a consequence of these postulates, one always has $P(A) + P(\overline{A}) = 1$ where $\overline{A} := \{\omega \in \Omega : \ w \notin A\}$, and so $P(\emptyset) = 0$.

Suppose that we now relax the demand for a single number expressing the probability of an event A, allowing assessment of uncertainty by an interval of numbers $[\underline{P}(A), \overline{P}(A)]$, with $\underline{P}(A)$ construed as the "lower probability" of A and $\overline{P}(A)$ as the "upper probability" of A. What properties should any uncertainty measures deserving of these names possess?

At the very least, we should have

$$\begin{aligned} 0 &\leq \underline{P}(A) \leq \overline{P}(A) \leq 1, \text{ for all events } A, \text{ with} \\ \underline{P}(\emptyset) &= \overline{P}(\emptyset) = 0 \text{ and } \underline{P}(\Omega) = \overline{P}(\Omega) = 1, \end{aligned} \tag{1}$$

as well as *monotonicity* of $\underline{P}$ and $\overline{P}$, i.e.,

$$\begin{aligned} A_1 \subseteq A_2 &\Rightarrow \underline{P}(A_1) \leq \underline{P}(A_2) \text{ and} \\ \overline{P}(A_1) &\leq \overline{P}(A_2) \end{aligned} \tag{2}$$

And if the bounding functions $\underline{P}$ and $\overline{P}$ are to be useful, there ought, of course, to exist at least one probability measure P satisfying

$$\underline{P}(A) \leq P(A) \leq \overline{P}(A) \quad \text{for all events } A. \tag{3}$$

When Ω is finite, checking that (3) holds for some P amounts to checking that certain linear inequalities have a solution, which is easily done by linear programming if Ω is not too large.

In what follows, therefore, we shall call functions $\underline{P}$ and $\overline{P}$ a *pair of lower and upper probability measures* if they satisfy (1), (2), and (3). It should be noted that some authors are more stringent in their use of these terms, requiring in addition the properties of *complementarity*, i.e.,

$$\underline{P}(A) + \overline{P}(\overline{A}) = 1 \quad \text{for all events } A, \tag{4}$$

superadditivity of $\underline{P}$, i.e.,

$$A_1 \cap A_2 = \emptyset \Rightarrow \underline{P}(A_1 \cup A_2) \geq \underline{P}(A_1) + \underline{P}(A_2) \tag{5}$$

and *subadditivity* of $\overline{P}$, i.e.

$$A_1 \cap A_2 = \emptyset \Rightarrow \overline{P}(A_1 \cup A_2) \leq \overline{P}(A_1) + \overline{P}(A_2). \tag{6}$$

The above three properties, are always possessed by the "tightest possible" lower and upper bounds defining the same class of probability measures P as (3). If Ω is not too large, one can upgrade a pair of upper and lower probability measures $\underline{P}$ and $\overline{P}$ to a pair $\underline{P}^{\#}$, $\overline{P}^{\#}$ satisfying (1)-(6), with $\underline{P}(E) \leq \underline{P}^{\#}(E) \leq \overline{P}^{\#}(E) \leq \overline{P}(E)$ for all events E. Using standard linear programming, one simply computes

$$\underline{P}^{\#}(E) = min\{P(E) : \; P \in \mathcal{P}(\underline{P}, \overline{P})\}, \quad \text{and} \tag{7}$$

where $\mathcal{P}(\underline{P}, \overline{P})$ is the closed, convex polyhedral set given by

$$\begin{aligned} \mathcal{P}(\underline{P}, \overline{P}) &:= \{P : P \text{ is a probability measure and} \\ \underline{P}(A) &\leq P(A) \leq \overline{P}(A) \text{ for all events } A\}, \end{aligned} \tag{8}$$

and then defines

$$\overline{P}^{\#}(E) = 1 - \underline{P}^{\#}(\overline{E}). \tag{9}$$

In general, one can of course avoid assessing values of both $\underline{P}$ and $\overline{P}$, assessing only values of $\underline{P}$ and *defining* $\overline{P}(E) = 1 - \underline{P}(\overline{E})$.

3 Choquet Expectation

Suppose that the random variable $X : \Omega \to \mathbb{R}$ takes on only a finite set of values, say, $x_1 < x_2 < \ldots < x_n$. If P is a probability measure on Ω, then the *expected value of X with respect to* P, denoted $\mathcal{E}_P(X)$, is given by the familiar formula

$$\mathcal{E}_P(X) = \sum_{i=1}^{n} x_i \; P(X = x_i), \tag{10}$$

where $P(X = x_i)$ is an abbreviation for $P(\{\omega \in \Omega : X(w) = x_i\})$.

By additivity of P we have, for $1 \leq i \leq n-1$, that $P(X = x_i) = P(X \geq x_i) - P(X \geq x_{i+1})$, so an equivalent (though slightly odd looking) formula for $\mathcal{E}_P(X)$ is given by

$$\begin{aligned} &\mathcal{E}_P(X) = \\ &\sum_{i=1}^{n-1} x_i\{P(X \geq x_i) - P(X \geq x_{i+1})\} + x_n P(X = x_n) \\ &= x_1 + \sum_{i=2}^{n} (x_i - x_{i-1}) P(X \geq x_i). \end{aligned} \tag{11}$$

Now if $\underline{P}$ and $\overline{P}$ are a pair of lower and upper probability measures on Ω, then, motivated by (11), we *define* $\mathcal{E}_{\underline{P}}(X)$ and $\mathcal{E}_{\overline{P}}(X)$ by the formulas

$$\mathcal{E}_{\underline{P}}(X) := x_1 + \sum_{i=2}^{n} (x_i - x_{i-1}) \underline{P}(X \geq x_i) \tag{12}$$

and

$$\mathcal{E}_{\overline{P}}(X) := x_1 + \sum_{i=2}^{n} (x_i - x_{i-1}) \overline{P}(X \geq x_i). \tag{13}$$

We call $\mathcal{E}_{\underline{P}}(X)$ and $\mathcal{E}_{\overline{P}}(X)$ the *Choquet expected values of X with respect to* $\underline{P}$ and $\overline{P}$. The above formulas are simply special cases of the general formula, valid for every random variable X,

$$\mathcal{E}_\alpha(X) = \int_0^\infty \alpha(X \geq x) dx - \int_{-\infty}^0 [1 - \alpha(X \geq x)] dx,$$

where $\alpha = \underline{P}, \overline{P}, P$ [2].

Since the quantities $x_i - x_{i-1}$ appearing in (12) and (13) are positive, it follows immediately that

$$P \in \mathcal{P}(\underline{P}, \overline{P}) \Rightarrow \mathcal{E}_{\overline{P}}(X) \leq \mathcal{E}_P(X) \leq \mathcal{E}_{\overline{P}}(X), \tag{14}$$

where $\mathcal{P}(\underline{P}, \overline{P})$ is defined by (8), and so, with

$$\begin{aligned} \underline{\mathcal{E}}(X) &:= min\{E_P(X) : P \in \mathcal{P}(\underline{P}, \overline{P})\}, \text{ and} \\ \overline{\mathcal{E}}(X) &:= max\{E_P(X) : P \in \mathcal{P}(\underline{P}, \overline{P})\}, \end{aligned} \tag{15}$$

it follows that

$$\mathcal{E}_{\underline{P}}(X) \leq \underline{\mathcal{E}}(X) \leq \overline{\mathcal{E}}(X) \leq \mathcal{E}_{\overline{P}}(X). \tag{16}$$

So the crucial quantities $\underline{\mathcal{E}}(X)$ and $\overline{\mathcal{E}}(X)$ may be conservatively approximated, respectively, by the easily computable quantities $\mathcal{E}_{\underline{P}}(X)$ and $\mathcal{E}_{\overline{P}}(X)$.

Note that the number of constraints defining $\mathcal{P}(\underline{P}, \overline{P})$ is exponential in the cardinality of Ω, so for "large" Ω the direct calculation of $\underline{\mathcal{E}}(X)$ and $\overline{\mathcal{E}}(X)$ may not be possible.

In certain cases, we are guaranteed to have $\mathcal{E}_{\underline{P}}(X) = \underline{\mathcal{E}}(X)$ and $\mathcal{E}_{\overline{P}}(X) = \overline{\mathcal{E}}(X)$. This always happens, for example, if the pair $\underline{P}$, $\overline{P}$ satisfy(1)-(4) and $\underline{P}$ satisfies the following stronger version of superadditivity, called 2-*monotonicity*:

$$\underline{P}(A_1 \cup A_2) \geq \underline{P}(A_1) + \underline{P}(A_2) - \underline{P}(A_1 \cap A_2). \tag{17}$$

(See [1] and [7].) Several common constructions of imprecise probabilities yield 2-monotone lower probability measures [4], [6], [8].

In the next section, we review a study in which the above ideas were applied to forecasting production costs of an electric utility.

4 Forecasting Production Costs

In a study of Thorp, McClure, and Fine [7], the 1990 production cost, C, of an actual, but unidentified, electric utility depends on the values of the uncertain quantities (i.e., random variables) E = energy demand (GWH), F = load factor, and R = average year to year coal price increase from 1981 to 1990 (%). Given specific values $E = e$, $F = f$, and $R = r$ of these random variables, a standard production costing algorithm can be used to calculate the corresponding production cost $C = C(e, f, r)$.

In this study the possible values of E, F, and R are given, respectively, by the sets Ω_E = { 50,000 GWH, 60,000 GWH, 70,000 GWH}, Ω_F = {0.635, 0.675, 0.725} and Ω_R = {1%, 2%, 4%}. To complete the construction of the conceptual apparatus of §3, the 27-element set Ω was defined by

$$\begin{aligned} \Omega &= \Omega_E \times \Omega_F \times \Omega_R \\ &:= \{(e, f, r) : e \in \Omega_E,\ f \in \Omega_F, \text{ and } r \in \Omega_R\}. \end{aligned} \tag{18}$$

The 1990 production cost is then a random variable $C: \ \Omega \to [0, \infty)$.

Next, the values of C were calculated for each of the 27 triples $(e, f, r) \in \Omega$ and arranged in increasing order, $c_1 < c_2 < \ldots < c_{27}$. Then, for each $i = 1, \ldots, 27$, the triples comprising each of the 27 events

$$\begin{aligned} A_i &:= \text{“}C \geq c_i\text{”} \\ &= \{(e, f, g) \in \Omega : C(e, f, g) \geq c_i\} \end{aligned} \tag{19}$$

were identified.

At this point, upper and lower probabilities $\underline{P}(A_i)$ and $\overline{P}(A_i)$ were assessed for $A_1, \ldots, A_{27}$. Then the Choquet expected values of C with respect to $\underline{P}$ and $\overline{P}$ were calculated by formulas (12) and (13), yielding in this case $\mathcal{E}_{\underline{P}}(C) = \$1.159\ B$ and $\mathcal{E}_{\overline{P}} = 1.438\ B$.

Thorp, McClure, and Fine assessed the required values of $\underline{P}$ and $\overline{P}$ by convening a panel of six experts from the planning department of the company. Each of the experts assessed probabilities over the separate sets $\Omega_E = \{50,000;\ 60,000;\ 70,000\}$, $\Omega_F = \{0.635,\ 0.675,\ 0.725\}$, and $\Omega_R = \{1, 2, 4\}$. Assuming independence, probabilities were multiplied to yield probability measures $P_1, \ldots, P_6$ on $\Omega = \Omega_E \times \Omega_F \times \Omega_R$. The lower and upper probability measures on Ω were constructed as the so-called *lower and upper envelopes of the family* $\{P_1, \ldots, P_6\}$, namely $\underline{P}(A_i) = min\{P_1(A_i), \ldots, P_6(A_i)\}$, and $\overline{P}(A_i) = max\{P_1(A_i), \ldots P_6(A_i)\}$, for $1 \leq i \leq 27$.

In [13] it is argued in detail that the family $\mathcal{P}(\underline{P}, \overline{P})$, with $\underline{P}$ and $\overline{P}$ constructed as above, does not represent the probability measures on Ω compatible with the experts' opinions. In particular, this family contains probability measures for which E, F, and R are not independent, and excludes other reasonable probability measures on Ω.

As an alternative to Choquet expectation in this particular case, we have explored bounding expected cost as follows: with $\Omega_E = \{e_1, e_2, e_3\}$, $\Omega_F = \{f_1, f_2, f_3\}$, and $\Omega_R = \{r_1, r_2, r_3\}$, the probability measures π on $\Omega = \Omega_E \times \Omega_F \times \Omega_R$ compatible with the experts' opinions are given by

$$\pi(e_i, f_j, r_k) = \epsilon_i \varphi_j \rho_k \tag{20}$$

where $\epsilon_1 + \epsilon_2 + \epsilon_3 = \varphi_1 + \varphi_2 + \varphi_3 = \rho_1 + \rho_2 + \rho_3 = 1$, and the partial sums of these variables are bounded by the minimum and maximum values assigned by the experts to the corresponding events in Ω_E, Ω_F, and Ω_R (e.g., $\epsilon_1 + \epsilon_2$ is bounded below by the smallest probability assigned by any expert to the event $\{e_1, e_2\}$ and above by the largest such probability).

The upper and lower bounds $\mathcal{E}^*(C)$ and $\mathcal{E}_*(C)$ are then calculated as the maximum and minimum of the nonlinear function

$$\mathcal{E}_\pi(C) = \sum_{\substack{\cancel{1 \leq i \leq j \leq k} \\ 1 \leq i, j, k \leq 3}} C(e_i, f_j, r_k) \epsilon_i \varphi_j \rho_k, \tag{21}$$

yielding $\mathcal{E}^*(C) = \$1.542\ B$ and $\mathcal{E}_*(C) = \$1.081\ B$, a somewhat wider pair of bounds than those calculated by Choquet expectation.

The concavity of $\mathcal{E}_\pi(C)$ and the special nature of the constraints makes the calculation of $\mathcal{E}_*(C)$ in this and considerably larger scale problems relatively simple. The calculation of $\mathcal{E}^*(C)$ is easy enough in this case, but may present difficulties in larger scale problems. We are currently investigating this issue. Of course, independence need not be assumed to apply the above method, for one can easily introduce variables representing conditional probabilities.

It should be noted that Choquet expectation, while inappropriate in the above problem, can be a very useful tool for bounding expected values, in particular when Ω is not a Cartesian product and upper and lower probabilities are directly assessed over events in Ω.

5 Conditionalization

Suppose that $\underline{P}$ and $\overline{P}$ are a pair of lower and upper probability measures on Ω. The probability measures P compatible with $\underline{P}$ and $\overline{P}$ are, we recall, just those $P \in \mathcal{P}(\underline{P}, \overline{P}) = \{P : \underline{P}(A) \leq P(A) \leq \overline{P}(A)$ for all events A$\}$. Now if some probability measure P is our model of "how uncertainties lie" in Ω and we are apprised of additional information with renders it certain that "the truth lies in E," for some subset $E \subseteq \Omega$ with $P(E) > 0$, it is customary to revise P by "conditionalization" to a new probability measure $P(\cdot|E)$, where $P(A|E) = P(A \cap E)/P(E)$ for all events A.

If, instead, we have P delineated only by $\underline{P}$ and $\overline{P}$, and we discover that E is certain, we have the problem of conditionalizing $\underline{P}$ and $\overline{P}$ (assume $\underline{P}(E) > 0$). A natural way to do this would be by the formulas

$$\underline{P}(A|E) = \min\{P(A|E) : P \in \mathcal{P}(\underline{P}, \overline{P})\} \tag{22}$$

and

$$\overline{P}(A|E) = \max\{P(A|E) : P \in \mathcal{P}(\underline{P}, \overline{P})\}. \tag{23}$$

That is, $\underline{P}(\cdot|E)$ and $\overline{P}(\cdot|E)$ are just the lower and upper envelopes of the family of all conditionalized probability measures $P(\cdot|E)$ as P runs through the set of all probability measures compatible with $\underline{P}$ and $\overline{P}$. As one would expect, it is in many cases impossible to compute $\underline{P}(A|E)$ and $\overline{P}(A|E)$ exactly. The difficulty is the very one encountered with respect to $\underline{\mathcal{E}}(X)$ and $\underline{\mathcal{E}}(X)$ in §3. In fact, the situation here is completely analogous to that of §3, for here we can find conservative approximations to $\underline{P}(A|E)$ and $\overline{P}(A|E)$ that are exact when $\underline{P}$ is two-monotone.

The approximations are easily derived. Let $P \in \mathcal{P}(\underline{P}, \overline{P})$. Since $P(E) = P(A \cap E) + P(\overline{A} \cap E)$ for every event A, one has

$$\begin{aligned} P(A|E) &= \frac{P(A \cap E)}{P(A \cap E) + P(\overline{A} \cap E)} \\ &\geq \frac{\underline{P}(A \cap E)}{\underline{P}(A \cap E) + P(\overline{A} \cap E)} \\ &\geq \frac{\underline{P}(A \cap E)}{\underline{P}(A \cap E) + \overline{P}(\overline{A} \cap E)}, \end{aligned} \tag{24}$$

where the first inequality holds because for fixed $c > 0$, $x/x+c$ is an increasing function of x for $x > 0$, and the second inequality holds because $P(\overline{A} \cap E)$ is replaced by the value $\overline{P}(\overline{A} \cap E) \geq P(\overline{A} \cap E)$. Similarly, one can show that

$$P(A|E) \leq \frac{\overline{P}(A \cap E)}{\overline{P}(A \cap E) + \underline{P}(\overline{A} \cap E)}. \tag{25}$$

From (22) - (25), it follows that

$$\begin{aligned} \frac{\underline{P}(A \cap E)}{\underline{P}(A \cap E) + \overline{P}(\overline{A} \cap E)} &\leq \underline{P}(A|E) \\ &\leq \overline{P}(A|E) \\ &\leq \frac{\overline{P}(A \cap E)}{\overline{P}(A \cap E) + \underline{P}(\overline{A} \cap E)}. \end{aligned} \tag{26}$$

Moreover, if $\underline{P}$ and $\overline{P}$ satisfy, in addition to the defining properties (1) - (3), the complementarity property (4), and if $\underline{P}$ is two-monotone (17), then the first and third inequalities in (26) are actually equalities. And in such a case, $\underline{P}(A|E) + \overline{P}(\overline{A}|E) = 1$, and $\underline{P}(\cdot|E)$ remains two-monotone [6].

Acknowledgement

This research was funded by Martin Marietta Energy Systems under Contract 11X-SN818V with the University of Tennessee.

References

[1] A. Chateauneuf and J. Jaffray, "Some characterizations of lower probabilities and other monotone capacities through the use of Möbius inversion," *Mathematical Social Sciences*, Vol. 17, pp. 263-283, 1984.

[2] P. Fishburn, *Nonlinear Preference and Utility Theory*, Johns Hopkins Press, 1988.

[3] E. Hurst, "A good integrated resource plan: guidelines for electric utilities and regulators," Oak Ridge National Laboratory ORNL/CON-354, 1992.

[4] V. Strassen, "Messfehler und information," *Z. Wahrscheinlichkeitstheorie*, Vol. 2, pp. 273-305, 1964.

[5] C. Sundberg and C. Wagner, "Characterizations of monotone and 2-monotone capacities," *Journal of Theoretical Probability*, Vol. 5, pp. 159-167, 1992.

[6] C. Sundberg and C. Wagner, "Generalized finite differences and Bayesian conditioning of Choquet capacities," *Advances in Applied Mathematics*, Vol. 13, pp. 262-272, 1992.

[7] Thorp, McClure, & Fine, "The use of expert opinion in forecasting production cost of an electric utility," *Proc. 1982; IEEE Internat Large Scale Systems Symp*, IEEE Press, New York, 1982.

[8] C. Wagner, "Consensus for belief functions and related uncertainty measures," *Theory and Decision*, Vol. 26, pp. 295-304, 1989.

[9] C. Wagner and B. Tonn, "Diagnostic conditionalization", *Mathematical Social Sciences*, Vol. 19, pp. 159-165, 1990.

[10] C. Wagner, "Generalizing Jeffrey conditionalization," *Uncertainty in Artificial Intelligence*, Vol. 92, Morgan Kaufmann, San Mateo, CA, 1992.

[11] C. Wagner and B. Tonn, "Constructing lower probabilities," *Uncertainty in Artifical Intelligence*, Vol. 93, Morgan Kaufmann, San Mateo, CA, 1993.

[12] C. Wagner, "Generalized probability kinematics," *Erkenntnis*, Vol. 36, pp. 245-257, 1992.

[13] C. Wagner, "Imprecise probability and integrated resource planning," *Research Report*, Dept. of Mathematics, University of Tennessee, 1994.

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

[15] P. Walley, and T. Fine, "Towards a frequentist theory of upper and lower probability," *Ann. Statist.*, Vol. 10, pp. 741-761, 1982.