How to Select A Model If We Know Probabilities with Interval Uncertainty

Vladik Kreinovich

Department of Computer Science
University of Texas at El Paso, 500 W. University
El Paso, Texas 79968, USA, vladik@utep.edu

Abstract

Purpose: When we know the probability of each model, a natural idea is to select the most probable model. However, in many practical situations, we do not know the exact values of these probabilities, we only know intervals that contain these values. In such situations, a natural idea is to select some probabilities from these intervals and to select a model with the largest selected probabilities. The purpose of this study is to decide how to most adequately select these probabilities.

Design/methodology/approach: We want the probability-selection method to preserve independence: If, according to the probability intervals, the two events were independent, then the selection of probabilities within the intervals should preserve this independence.

Findings: We describe all techniques for decision making under interval uncertainty about probabilities that are consistent with independence. We prove that these techniques form a 1-parametric family, a family that has already been successfully used in such decision problems.

Originality/value: We provide a theoretical explanation of an empirically successful technique for decision making under interval uncertainty about probabilities. This explanation is based on the natural idea that the method for selecting probabilities from the corresponding intervals should preserve independence.

Keywords: Decision making under uncertainty; Interval uncertainty about probabilities; Independent events; Maximum likelihood approach.

1 Formulation of the Problem

Need for indirect measurements and data processing. In many practical situations, we are interested in the quantity y that is difficult – or even impossible to measure directly. For example, we may be interested in tomorrow's temperature or in next year's Gross Domestoc Product (GDP). Since we cannot measure the quantity y directly, we need to measure it *indirectly*, i.e.:

- find easier-to-measure quantities $x_1, \ldots, x_n$ which are related to y by a known dependence $y = f(x_1, \ldots, x_n)$,
- measure the values of these quantities, resulting in measurement results $\widetilde{x}_1, \ldots, \widetilde{x}_n$, and
- compute the desired estimate $\widetilde{y} = f(\widetilde{x}_1, \ldots, \widetilde{x}_n)$ by applying the algorithm f to the results $\widetilde{x}_i$ of measuring x_i.

Computing $\widetilde{y}$ is an important particular case of *data processing.*

Need to find a model. In many practical situations, we know the function $f(x_1, \ldots, x_n)$. For example, in celestial mechanics, we know how the future location y of a celestial body depends on the current location and velocity of this and other bodies. However, in many other practical situations, we do nor know this dependence. In such cases, we need to determine this dependence from the experiments and/or observations. Specifically;

- in several (K) cases, we know both the values $x_i^{(k)}$ of the inputs x_i and the value $y^{(k)}$ of the desired quantity y, and
- we need to find the dependence $f(x_1, \ldots, x_n)$ that is consistent with all these observations, i.e., for which, for all k from 1 to K, we have

$$y^{(k)} = f\left(x_1^{(k)}, \ldots, x_n^{(k)}\right). \quad (1)$$

Terminological comment.

- The resulting function $y = f(x_1, \ldots, x_n)$ serves as a *model* of the corresponding situation.
- In statistics, the problem of finding a model is known as *regression.*
- In computer science, the same problem – when solved by an algorithm – is known was *machine learning.*
- In this paper, we use the word "model" in the general scientific sense – as a description of a real-life process, i.e., in this case, as a function $f(x_1, \ldots, x_n)$ that estimates the desired quantity y. To avoid possible confusion, it should be mentioned that in statistics, sometimes, a "model" means a *family* of such functions – e.g., all linear functions or all linear functions that depend only on the first k variables $x_1, \ldots, x_k$.

Need to select a model.

- To describe a general function $f(x_1, \ldots, x_n)$, we need to describe infinitely many parameters – e.g., the values of the function at all the tuples $(x_1, \ldots, x_n)$ for which all the values x_i are rational.
- Our on;y requirement on possible functions is to satisfy K equations (1).

Here, the number of parameters much larger than the number of equations. Thus, there are, in general, many different functions that fits all the observations.

We therefore need to select one of these functions, i.e., we need to select a model.

How a model is selected now: case when we know probabilities. In some cases, we know the probabilities p_i of different models. In this case, a reasonable idea is to select the *most probable* model, i.e., the model whose probability is the largest: $p_i = \max_j p_j$.

Such a selection is, e.g., one of the main ideas behind the maximum likelihood approach to model selection; see, e.g., (Sheskin, 2011). In this method, usually, we maximize the probability p by solving the equivalent problem of minimizing the quantity

$$L \stackrel{\text{def}}{=} -\ln(p)\,.$$

Comment. It should be mentioned that, strictly speaking, likelihood is not the probability of a model, it is the probability of the data according to this model. To come up with the probability of the model, we need to use Bayesian approach. In this approach, if we assume that a priori all models are equally probable – i.e., that prior distribution is uniform – then likelihood becomes proportional to the probability of the model, so that maximizing likelihood is equivalent to maximizing the model's probability.

What if we only have partial information about probabilities: description of the situation. Often, we only have partial information about the probabilities. For example, instead of the exact values p_i of each probability, we only know the lower bound $\underline{p}_i$ and the upper bound $\overline{p}_i$: $\underline{p}_i \le p_i \le \overline{p}_i$.

In this case, the only information that we have about the probability p_i is that this probability is contained on the interval $\left[\underline{p}_i, \overline{p}_i\right]$. Thus, this situation is known as *interval uncertainty*.

How to make a decision under such interval uncertainty: a natural idea. in situations with interval uncertainty, it is desirable to apply well-traditional probability-based decision making technique. To do this, we need to select, within each of the intervals $\left[\underline{p}_i, \overline{p}_i\right]$, one of the values p_i, and then select the model with the largest value of this selected probability p_i.

Resulting challenge. How do we select a value p_i in each interval? There are many different ways to select, which one should we choose?

What we do in this paper. In this paper, we show that a natural condition on the selection of the probability values from the corresponding intervals uniquely determine a 1-parametric family of such selections – the only selections that satisfy this natural condition.

2 Main Result

Natural condition: informal description. We want to find a mapping that assigns, to each interval of probability values, a number from this interval. It is desirable to select this mapping so that it preserves important properties of the situation.

In probabilistic techniques, one of the most important notions is the notion of independence. It is therefore reasonable to require that the desired intervals-to-numbers mapping satisfy the following condition: If the two events were independent, then this mapping should preserve this independence.

Let us formalize this natural condition. If two events with probabilities p_1 and p_2 are independent, then the probability of them occurring at same time is equal to the product $p_1 \cdot p_2$ of the corresponding probabilities. If for each of these events, we only know the interval $\left[\underline{p}_i, \overline{p}_i\right]$ of possible values of its probability, then possible values of the probability that both events occur is equal to the set of possible values

$$\left\{p_1 \cdot p_2 : p_1 \in \left[\underline{p}_1, \overline{p}_1\right] \text{ and } p_2 \in \left[\underline{p}_2, \overline{p}_2\right]\right\}.$$

One can easily check that this set is equal to the interval $\left[\underline{p}_1 \cdot \underline{p}_2, \overline{p}_1 \cdot \overline{p}_2\right]$; see, e.g., (Jaulin at al., 2012), (Kubica, 2019), (Mayer, 2017), (Moore, Kearfott, and Cloud, 2009). Indeed, for non-negative values p_i, the product function

$p_1, p_2 \mapsto p_1 \cdot p_2$ is (non-strictly) increasing with respect to each of its variables. Thus:

- the smallest possible value of this function when $p_i \in \left[\underline{p}_i, \overline{p}_i\right]$ is attained when both inputs are the smallest possible, i.e., when $p_i = \underline{p}_i$ for both i, and
- the largest possible value of this function when $p_i \in \left[\underline{p}_i, \overline{p}_i\right]$ is attained when both inputs are the largest possible, i.e., when $p_i = \overline{p}_i$ for both i.

Thus, we arrive at the following definition.

Definition. *We say that a mapping f that maps each subinterval $\left[\underline{p}, \overline{p}\right]$ of the interval $[0, 1]$ into a number $f\left(\underline{p}, \overline{p}\right)$ from this interval is* natural *if it satisfies the following condition: for all values $\underline{p}_1 \le \overline{p}_1$ and $\underline{p}_2 \le \overline{p}_2$, we have*

$$f\left(\underline{p}_1 \cdot \underline{p}_2, \overline{p}_1 \cdot \overline{p}_2\right) = f\left(\underline{p}_1, \overline{p}_1\right) \cdot f\left(\underline{p}_2, \overline{p}_2\right).$$

Proposition. *A mapping is natural if and only if, for some $\alpha \in [0, 1]$, it has the form*

$$f\left(\underline{p}, \overline{p}\right) = \underline{p}^{\alpha} \cdot \overline{p}^{1-\alpha}.$$

Discussion. The function $L = -\ln(p)$ is decreasing with respect to p. Thus, when $p \in \left[\underline{p}, \overline{p}\right]$, then:

- the smallest value $\underline{L}$ of $L = -\ln(p)$ is attained when p is the largest, i.e., when $p = \overline{p}$:
$$\underline{L} = -\ln\left(\overline{p}\right);$$
- the largest value $\overline{L}$ of L is attained when p is the smallest, i.e., when $p = \underline{p}$:
$$\overline{L} = -\ln\left(\underline{p}\right).$$

For the values $L = -\ln(p)$, $\underline{L} = -\ln\left(\overline{p}\right)$, and $\overline{L} = -\ln\left(\underline{p}\right)$, the above formula takes the form $L = \alpha \cdot \overline{L} + (1-\alpha) \cdot \underline{L}$. Interestingly, this is exactly

Hurwicz optimism-pessimism criterion that is used for decision making under interval uncertainty; see, e.g., (Hurwicz, 1951), (Kreinovich, 2014), (Luce and Raiffa, 1989).

This model selection has been successfully used; see, e.g., (Denoeux, 2023).

Proof of the Proposition.

1°. It is easy to prove that the above formula leads to a natural mapping. So, to complete the proof, it is sufficient to prove that every natural mapping has this form.

Let $f\left(\underline{p}, \overline{p}\right)$ be a natural mapping. Let is prove that it has the desired form.

2°. For each p, by definition of a natural mapping, the value $f(p,p)$ belongs to the interval $[p,p]$ and is, thus, equal to p. In particular, for $p=0$, we get

$$f(0,0)=0.$$

3°. Let us first take $\underline{p}_1=\underline{p}_1=0$ and $\overline{p}_2=\overline{p}_2=1$. In this case, the naturalness condition implies that $f(0,1)\cdot f(0,1)=f(0,1)$. Thus, either $f(0,1)=1$ or $f(0,1)=0$. Let us consider these two possible cases one by one.

4°. Let us first consider the case when $f(0,1)=1$.

4.1°. In this case, for every $a\in[0,1]$, for $\underline{p}_1=0$, $\overline{p}_1=1$, $\underline{p}_2=1$, and $\overline{p}_2=1$, we get $f(0,1)\cdot f(a,1)=f(0,1)$. Since $f(0,1)=1$, this means that $f(a,1)=1$ for all a.

4.2°. Now, for all possible $\underline{p}\le\overline{p}$ for which $\overline{p}>0$, naturalness leads to

$$f\left(\underline{p},\overline{p}\right)=f\left(\overline{p},\overline{p}\right)\cdot f\left(\underline{p}/\overline{p},1\right).$$

As we have proven in Section 4.1 of this proof, the second factor $f\left(\underline{p}/\overline{p},1\right)$ is equal to 1. The first factor $f\left(\overline{p},\overline{p}\right)$ is, by Part 2 of this proof, equal to $\overline{p}$.

So, for all cases when $\overline{p}>0$, we have $f\left(\underline{p},\overline{p}\right)=\overline{p}$.

4.3°. For $\overline{p} = 0$, the formula $f\left(\underline{p}, \overline{p}\right) = \overline{p}$ is also true – by Part 2 of this proof. Thus, this formula holds for all $\underline{p} \le \overline{p}$. This corresponds to $\alpha = 0$.

5°. Let us now consider the case when $f(0,1) = 0$.

In this case, naturalness implies that, for all $\overline{p}$, we have

$$f(0, \overline{p}) = f(0,1) \cdot f(0, \overline{p})$$

and hence $f(0, \overline{p}) = 0$. Let us now consider intervals for which $\underline{p} > 0$.

5.1°. Let us first consider the values $f(a,1)$ corresponding to $a > 0$. When $a < b$, then we have $f(a,1) = f(a/b,1) \cdot f(b,1)$. Since $f(a/b,1)$ is a probability, it is smaller than or equal to 1, thus, $f(a,1) \le f(b,1)$, i.e., $f(a,1)$ is a non-strictly increasing function of a.

5.2°. Each value $a > 0$ can be represented as $\exp(-A)$ for $A = -\ln(a)$. By definition of the natural mapping, each such value $f(a,1)$ for $a > 0$ is greater than or equal to $a > 0$ and thus, $f(a,1) > 0$. So, we can take logarithm of these values as well. Let us denote $F(A) \stackrel{\text{def}}{=} -\ln(f(\exp(-A),1))$. Probabilities $f(\exp(-A),1)$ are smaller than or equal to 1, so

$$\ln(f(\exp(-A),1)) \le 1 = 0$$

and thus, for $F(A) \stackrel{\text{def}}{=} -\ln(f(\exp(-A),1))$, we always have $F(A) \ge 0$. In particular, $F(1) \ge 0$.

5.3°. Let us prove that $F(A)$ is a (non-strictly) increasing function.

Indeed, if $A < B$, then $-A > -B$. Since $\exp(x)$ is an increasing function, we get $\exp(-A) > \exp(-B)$. Since $f(a,1)$ is a non-strictly increasing function of a, we conclude that $f(\exp(-A),1) \ge f(\exp(-B),1)$. Since $\ln(x)$ is an increasing function, we conclude that

$$\ln(f(\exp(-A),1)) \ge \ln(f(\exp(-B),1)).$$

Multiplying both sides by -1, we get

$$-\ln(f(\exp(-A),1)) \le -\ln(f(\exp(-B),1)),$$

i.e., $F(A) \leq F(B)$. The statement is proven.

5.4°. For values $f(a,1)$, naturalness implies that $f(a \cdot b, 1) = f(a,1) \cdot f(b,1)$. For $a = \exp(-A)$ and $b = \exp(-B)$, we have $a \cdot b = \exp(-(A+B))$, thus,

$$f(\exp(-(A+B)), 1) = f(\exp(-A), 1) \cdot f(\exp(-B), 1).$$

By taking negative logarithms of both sides, we get

$$F(A+B) = F(A) + F(B). \quad (2)$$

5.5°. For every integer m, the formula (2) implies that

$$F(m \cdot A) = F(A + \ldots + A \ (m \text{ times})) =$$

$$F(A) + \ldots + F(A) \ (m \text{ times}) = m \cdot F(A). \quad (3)$$

In particular, for $m = n$ and $A = 1/n$, we get $F(1) = n \cdot F(1/n)$, hence

$$F(1/n) = (1/n) \cdot F(1). \quad (4)$$

For a general m and $A = 1/n$, we get $F(m/n) = m \cdot F(1/n)$. Due to (4), we get $F(m/n) = (m/n) \cdot F(1)$, i.e., $F(r) = r \cdot F(1)$ for all rational number r.

5.6°. For every real number x and for every positive integer n, we can take, as m_n, an integer part of $n \cdot x$, so that $m_n \leq n \cdot x < m_n + 1$. By dividing all parts of this inequality by n, we get $m_n/n \leq x < (m_n + 1)/n$. In the limit $n \to \infty$, we get $m_n/n \to x$ and $(m_n + 1)/n \to x$.

By Part 5.3 of this proof, the function $F(A)$ is non-strictly increasing, thus $F(m_n/n) \leq F(x) \leq F((m_n + 1)/n)$. Due to Part 5.5, this means that

$$(m_n/n) \cdot F(1) \leq F(x) \leq ((m_n + 1)/n) \cdot F(1).$$

In the limit $n \to \infty$, we have

$$(m_n/n) \cdot F(1) \to x \cdot F(1) \text{ and } ((m_n + 1)/n) \cdot F(1) \to x \cdot F(1).$$

Thus, in the limit, we get $x \cdot F(1) \le F(x) \le x \cdot F(1)$, i.e., $F(x) = x \cdot F(1)$.

5.7°. By definition, $F(A) = -\ln(f(\exp(-A), 1))$, thus,

$$f(\exp(-A), 1) = \exp(-F(A)) = \exp(-A \cdot F(1)).$$

Substituting $A = -\ln(a)$ into this expression, we get

$$f(a, 1) = e^{\ln(a) \cdot F(1)} = \left(e^{\ln(a)}\right)^{F(1)} = a^{F(1)}.$$

The condition that $f(a, 1) \ge a$ implies that $F(1) \le 1$, thus $F(1) \in [0, 1]$.

5.8°. For every pair $0 < \underline{p} \le \overline{p}$, naturalness implies that

$$f\left(\underline{p}, \overline{p}\right) = f(\overline{p}, \overline{p}) \cdot f\left(\underline{p}/\overline{p}\right).$$

By Part 2 of this proof, the first factor in this product is equal to $\overline{p}$. Due to Part 5.7, we get the expression for the second factor, thus we get

$$f\left(\underline{p}, \overline{p}\right) = \overline{p} \cdot \left(\underline{p}/\overline{p}\right)^{F(1)} = \underline{p}^{F(1)} \cdot \overline{p}^{1-F(1)}.$$

This is exactly the desired formula, for $\alpha = F(1)$ – limited to the case when $\underline{p} > 0$. Then:

- If $\alpha = 0$, we get the case considered in Part 4 of this proof.
- For $\alpha > 0$ and $\underline{p} = 0$, we have $0^\alpha \cdot \overline{p}^{1-\alpha} = 0$, and $f(0, \overline{p}) = 0$ by Part 5, so the desired equality holds for all $\underline{p} \le \overline{p}$.

The proposition is proven.

Acknowledgments

The author is thankful to Thierry Denœux, Hung T. Nguyen, and Nguyen Ngoc Thach, for their encouragement and advice.

References

Th. Denœux, "Quantifying prediction uncertainty in regression using random fuzzy sets: the ENNreg model", *IEEE Transactions on Fuzzy Systems*, 2023, to appear.

L. Hurwicz, *Optimality Criteria for Decision Making Under Ignorance*, Cowles Commission Discussion Paper, Statistics, No. 370, 1951.

L. Jaulin, M. Kiefer, O. Didrit, and E. Walter, *Applied Interval Analysis, with Examples in Parameter and State Estimation, Robust Control, and Robotics*, Springer, London, 2012.

V. Kreinovich, "Decision making under interval uncertainty (and beyond)", In: P. Guo and W. Pedrycz (eds.), *Human-Centric Decision-Making Models for Social Sciences*, Springer Verlag, 2014, pp. 163–193.

B. J. Kubica, *Interval Methods for Solving Nonlinear Constraint Satisfaction, Optimization, and Similar Problems: from Inequalities Systems to Game Solutions*, Springer, Cham, Switzerland, 2019.

R. D. Luce and R. Raiffa, *Games and Decisions: Introduction and Critical Survey*, Dover, New York, 1989.

G. Mayer, *Interval Analysis and Automatic Result Verification*, de Gruyter, Berlin, 2017.

R. E. Moore, R. B. Kearfott, and M. J. Cloud, *Introduction to Interval Analysis*, SIAM, Philadelphia, 2009.

D. J. Sheskin, *Handbook of Parametric and Nonparametric Statistical Procedures*, Chapman and Hall/CRC, Boca Raton, Florida, 2011.