Why People Tend to Overestimate Joint Probabilities ⋆

Olga Kosheleva[0000−0003−2587−4209] and
Vladik Kreinovich[0000−0002−1244−1650]

University of Texas at El Paso, El Paso, TX 79968, USA
{olgak,vladik}@utep.edu
https://www.cs.utep.edu/vladik/olgavita.html
https://www.cs.utep.edu/vladik/longvita.html

Abstract. It is known that, in general, people overestimate the probabilities of joint events. In this paper, we provide an explanation for this phenomenon – as explanation based on Laplace Indeterminacy Principle and Maximum Entropy approach.

Keywords: Subjective probability · Interval uncertainty · Maximum Entropy approach · Laplace Indeterminacy Principle · Fuzzy logic

1 Formulation of the Problem

1.1 Description of the situation

In many practical situations, we need to estimate the probability that two events A and B both happen – i.e., in other words, we need to estimate the joint probability $P(A\,\&\,B)$. Often, the only information that we have about two events A and B are their probabilities $a = P(A)$ and $b = P(B)$, and we have no information about the relation between these two events.

In this case, what should be a reasonable estimate for $P(A\,\&\,B)$?

1.2 A natural solution to such situations

This is an example of a situation when we need to make a decision under uncertainty, i.e., in this case, when we do not have full information about the probabilities. Such situations are ubiquitous, and in probability theory, there is

⋆ This work was supported in part by the National Science Foundation grants 1623190 (A Model of Change for Preparing a New Generation for Professional Practice in Computer Science), and HRD-1834620 and HRD-2034030 (CAHSI Includes), and by the AT&T Fellowship in Information Technology.
It was also supported by the program of the development of the Scientific-Educational Mathematical Center of Volga Federal District No. 075-02-2020-1478, and by a grant from the Hungarian National Research, Development and Innovation Office (NRDI).

a natural general approach to solving these problems. Its main idea is that different possible probability distributions correspond, in general, to different amount of uncertainty. A natural measure of uncertainty is the average number of binary ("yes"-"no") questions that we need to ask to uniquely determine the situation. It is known that in situations when we have n alternatives with probabilities $p_1, \ldots, p_n$, then this average number of questions is equal to Shannon's entropy $S = -\sum\limits_{i=1}^{n} p_i \cdot \log_2(p_i)$; see, e.g., [15].

It is reasonable not to pretend that we have less uncertainty than we do. For example, suppose that we know only that we have n alternatives, and we do not have any information about the probabilities of these alternatives. It may be that the probability of the first alternative is 1 and the probabilities of all other alternatives are 0s. In this case, we have no uncertainty at all – so the entropy is 0. But if we select this distribution, we will be inserting certainty where there is a lot of uncertainty – e.g., we can have a probability distribution in which $p_1 = \ldots = p_n = 1/n$, in which case $S = \log_2(n)$.

To avoid artificially inserting certainty, it is reasonable to select, among all possible distributions, the one with the largest possible value of entropy. This approach is known as the *Maximum Entropy approach*; see, e.g., [7].

In particular, for the case when we have n alternatives, and we have no information about their probabilities, Maximum Entropy approach implies that we select the distribution in which all these alternatives have the same probability $1/n$. This makes perfect sense: since we have no reason to prefer one of the alternatives, it makes sense to assign the same probability to all of them. This idea was first explicitly formulated by Laplace and is known as *Laplace Indeterminacy Principle.*

1.3 What if we apply this approach to our situation

In our situation, we have fours possible alternatives: $A \,\&\, B$, $A \,\&\, \neg B$, $\neg A \,\&\, B$, and $\neg A \,\&\, \neg B$. Once we select the probability $p = P(A \,\&\, B)$, we can find all other probabilities:

$$P(A \,\&\, \neg B) = P(A) - P(A \,\&\, B) = a - p,$$

$$P(\neg A \,\&\, B) = P(B) - P(A \,\&\, B) = b - p, \text{ and}$$

$$P(\neg A \,\&\, \neg B) = 1 - P(A \,\&\, B) - P(A \,\&\, \neg B) - P(\neg A \,\&\, B) =$$

$$1 - p - (a - p) - (b - p) = 1 + p - a - b.$$

In this case, the entropy has the form

$$-p \cdot \log_2(p) - (a-p) \cdot \log_2(a-p) - (b-p) \cdot \log_2(b-p) -$$

$$(1 + p - a - b) \cdot \log_2(1 + p - a - b). \quad (1)$$

Here, for every x, we have

$$\log_2(x) = \frac{\ln(x)}{\ln(2)}.$$

Maximizing the expression (1) is equivalent to maximizing the same expression multiplied by $\ln(3)$, i.e., the expression

$$-p\cdot\ln(p)-(a-p)\cdot\ln(a-p)-(b-p)\cdot\ln(b-p)-(1+p-a-b)\cdot\ln(1+p-a-b). \quad (2)$$

Differentiating this expression with respect to p and equating the derivative to 0, we conclude that

$$-1-\ln(p)+1+\ln(a-p)+1+\ln(b-p)-1-\ln(1+p-a-b)=0,$$

i.e.,

$$-\ln(p)+\ln(a-p)+\ln(b-p)-\ln(1+p-a-b)=0.$$

Applying the exponential function to both sides and taking into account that $\exp(0)=1$, $\exp(\ln(x))=x$, $\exp(a+b)=a+b$, and $\exp(a-b)=a/b$, we conclude that

$$\frac{p\cdot(1+p-a-b)}{(a-p)\cdot(b-p)}=1,$$

i.e., equivalently, that

$$p\cdot(1+p-a-b)=(a-p)\cdot(b-p).$$

Opening parentheses, we get

$$p+p^2-a\cdot p-b\cdot p=a\cdot b-a\cdot p-b\cdot p+p^2.$$

Canceling equal terms in both sides, we conclude that

$$p=a\cdot b.$$

So, in our situation, a reasonable estimate for the joint probability $P(A\,\&\,B)$ is the product $P(A)\cdot P(B)$, corresponding to the case when the events A and B are independent.

1.4 How do people actually estimate the joint probability

Empirical data shows that in situations when people know the probabilities $P(A)$ and $P(B)$ of individual events A and B – and they have no additional information – they often overestimate the probability $P(A\,\&\,B)$ of a joint event $A\,\&\,B$; see, e.g., [1, 17, 22, 23]. This happens both when we explicitly ask them to estimate the correspondiong probabilities, and when we extract their estimated probabilities from their preferences.

In other words, they usually provide an estimate which is larger than the product $P(A)\cdot P(B)$.

1.5 Formulation of the problem

How can we explain this phenomenon?

1.6 Important comment

In most of the above-cited experiments, people were given:

- either situations in which both A and B are somewhat rare events with non-zero probabilities, i.e., when $0 < a < 0.5$ and $0 < b < 0.5$,
- or situations in which both A and B are rather frequent events with some uncertainty, i.e., when $0.5 < a < 1$ and $0.5 < b < 1$.

2 Our Explanation

2.1 Main idea behind this explanation

In situations when we know the probabilities $a = P(A)$ and $b = P(B)$ of two events A and B, what are the possible values of the joint probability $p = P(A \,\&\, B)$? It is known that the set of such possible values is determined by the so-called Frechet inequality (see, e.g., [20]):

$$\max(a + b - 1, 0) \le p \le \min(a, b).$$

In other words, possible values of the joint probability p form an interval

$$[\max(a + b - 1, 0), \min(a, b)].$$

This is an example of an *interval uncertainty*; see, e.g., [6, 11, 13].

All values from this interval are possible, and we have no reason to conclude that some values are more probable than others. It is therefore reasonable to conclude that all these values are equally probable, i.e., that we have a uniform distribution on this interval. As we have mentioned earlier, such a conclusion – corresponding to Laplace Indeterminacy Principle – also follows from the Maximum Entropy approach.

According to decision theory (see, e.g., [4, 5, 9, 10, 14, 15, 19]), a rational person should make decisions based on the expected value of the utility. In case of a binary decision, this simply means using the expected (mean) value of the unknown probability p. For the uniform distribution on an interval, the mean value is the midpoint of this interval, i.e., the value

$$p = \frac{\max(a + b - 1, 0) + \min(a, b)}{2}. \quad (3)$$

2.2 This idea explains the observed overestimation: case when $0 < a < 0.5$ and $0 < b < 0.5$

Let us show that in both above cases, our main idea explains the observed overestimation of joint probabilities.

Let us start with the case when $0 < a < 0.5$ and $0 < b < 0.5$. In this case, $a + b < 1$, so $\max(a + b - 1, 0) = 0$, and the formula (3) takes the form

$$p = \frac{\min(a, b)}{2}. \quad (4)$$

There are two possible cases here: $a \le b$ and $b \le a$. If $b \le a$, then we can simply rename A and B as, correspondingly, B and A. Thus, without losing generality, we can assume that $a \le b$. In this case, the formula (4) implies that $p = a/2$. We need ro show that $a \cdot b < p = a/2$. Indeed, in this case, $b < 0.5$. Multiplying both sides of this inequality by $a > 0$, we get the desired inequality $a \cdot b < p/2$.

2.3 This idea explains the observed overestimation: case when $0.5 < a < 1$ and $0.5 < b < 1$

Let us now consider the case when $0.5 < a < 1$ and $0.5 < b < 1$. In this case, we $a+b > 1$, so $\max(a+b-1, 0) = a+b-1$. Similarly to the previous case, without losing generality, we can assume that $a \le b$ and thus, $\min(a, b) = a$. Then, the formula (3) takes the form

$$p = \frac{a+b-1+a}{2} = \frac{2a+b-1}{2}. \tag{5}$$

The desired inequality

$$a \cdot b < \frac{2a+b-1}{2} \tag{6}$$

is equivalent to

$$2 \cdot a \cdot b < 2a + b - 1,$$

i.e., by moving terms between sides, to

$$1 - b < 2a - 2a \cdot b = 2a \cdot (1-b). \tag{7}$$

Since $0.5 < a$, we have $1 < 2a$. Multiplying both sides of this inequality by $1-b > 0$, we get the inequality (7) – and thus, the desired inequality (6) – which is equivalent to (7).

So, in both cases, the reasonable estimate (3) is larger than the product $a \cdot b$ – which explains the observed overestimation of joint probabilities.

2.4 What about the general case?

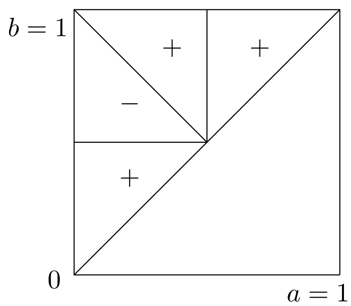
Without losing generality, we can assume that $a \le b$. We have considered cases when both a and b are smaller than 0.5 and when both a and b are larger than 0.5. So, the only remaining case is when $a \le 0.5 \le b$. We can have two subcases:

- the subcase when $a + b \le 1$, and
- the subcase when $a + b > 1$.

In the first subcase, we have $\max(a+b-1, 0) = 0$ and thus, $p = a/2$. In this case, since $0.5 \le b$, we have $a/2 \le a \cdot b$. So, we have either the exact estimate, or – when $a > 0$ and $b > 0.5$ – an *underestimation*. For example, when $a = 0.4$ and $b = 0.6$, the formula (3) leads to $p = 0.2 < 0.4 \cdot 0.6 = 0.24$.

In the second subcase, we have $\max(a + b - 1, 0) = a + b - 1$, and thus, similarly to the case when both a and b are larger than 0.5, we conclude that the value (3) is either the exact estimate – when $a = 0$ or when $b = 0.5$ – or an overestimation.

Let us describe, in the areas where $a \le b$, subareas of overestimation (+) and underestimation (−):



We have four equal-size triangles. In three out of four of them we have overestimation. This explains why in most cases, people overestimate joint probabilities.

3 Discussion

3.1 Is there an inconsistency here?

Before we discuss possible consequences of our explanation, we need to first clarify the situation, since what we described may sound fishy.

- In the first section of this paper, we used Maximum Entropy approach to conclude that, when we only know the probabilities $a = P(A)$ and $b = P(B)$, then the best estimate for $P(A \,\&\, B)$ must be the product $a \cdot b$.
- However, in the previous section, we used the same Maximum Entropy approach to come up with a completely different formula.

At first glance, this may sound like an inconsistency, but it is not.

It is well known that the same Maximum Entropy approach can lead to different answers – depending on how we formulate the problem. Let us give a simple example.

- Suppose that all we know is that a quantity x – e.g., the standard deviation – is somewhere on the interval $[0, 1]$. In this case, as we have mentioned earlier, the Maximum Entropy approach recommends that we assume that all the values from this interval are equally probable – i.e., that we have a uniform distribution on this interval.

– On the other hand, if all we know about x is that it is somewhere on the interval $[0,1]$, then all we know about $y = x^2$ – e.g., about variance – is that it is somewhere in the interval $[0,1]$. If we apply the same Maximum Entropy approach to the distribution of y, then we can conclude that y is also uniformly distributed on the interval $[0,1]$.

However, if x is uniformly distributed, then the distribution of $y = x^2$ is *not* uniform, so these conclusions are indeed different.

3.2 What if we have three of more events?

Suppose now that we know the probabilities $a_i = P(A_i)$ of n different events $A_1, \ldots, A_n$, and we want to estimate the joint probability $p = P(A_1 \,\&\, \ldots \,\&\, A_n)$. In this case, the only thing we know about p is that it belongs to the interval

$$[\max(a_1 + \ldots + a_n - (n-1), 0), \min(a_1, \ldots, a_n)],$$

so the same logic as for the above case of two events leads us to the conclusion that a reasonable estimate would be the midpoint of this interval, i.e., the value

$$\frac{\max(a_1 + \ldots + a_n - (n-1), 0) + \min(a_1, \ldots, a_n)}{2}. \tag{8}$$

It is important to mention that, e.g., for $n = 3$, this estimate is different from what we would have obtained if we:

– first use the two-event formula to estimate the probability of $A_1 \,\&\, A_2$, and
– then apply the same two-event formula to the events $A_1 \,\&\, A_2$ and A_3.

For example, if $a_1 = a_2 = a_3 = 0.6$, then:

– On the one hand, the formula (8) leads to

$$\frac{\max(1.8 - 2, 0) + 0.6}{2} = 0.3.$$

– On the other hand, our estimate for $P(A_1 \,\&\, A_2)$ is

$$\frac{\max(1.2 - 1, 0) + 0.6}{2} = \frac{0.2 + 0.6}{2} = 0.4;$$

using the formula (3) to combine 0.4 and 0.6, we get

$$\frac{\max(0.4 + 0.6 - 1, 0 + \min(0.4, 0.6)}{2} = \frac{0.4}{2} = 0.2,$$

which is indeed different from 0.3.

3.3 Computational conclusion

As we have mentioned, there are two reasonable way to apply Maximum Entropy approach to our situation, that lead to two different formulas: the product formula $a \cdot b$, and a different formula (3). As we have also mentioned, people, in general, overestimate the joint probabilities – i.e., produce estimates which are larger than $a \cdot b$.

So, if we want to adequately describe human reasoning, we need to use general "and"-combination rules which leads to values larger than $a \cdot b$. Such rules – under the name of t-norms – are typical is fuzzy logic; see, e.g., [2, 8, 12, 16, 18, 24]. So, we arrive at one more argument that fuzzy techniques are necessary if we want to adequately describe human reasoning – and an adequate description of such reasoning is one of the objectives of AI.

3.4 Physical conclusions

In physical terns, the fact that the joint probability is, in general, larger than the probability $a \cdot b$ corresponding to independent events means that there is, in general, a correlation between many events. In plain English, this means that everything in the world is interconnected – when formulated in these terms, this becomes a truism: everything in the world *is* interconnected; see, e.g., [3, 21].

References

1. Bar-Hillel, M.: On the subjective probability of compound events. Organizational Behavior and Human Performance **9**(4), 396–406 (1973)
2. Belohlavek, R., Dauben, J. W., Klir, G. J.: Fuzzy Logic and Mathematics: A Historical Perspective. Oxford University Press, New York (2017)
3. Feynman, R., Leighton, R., Sands, M.: The Feynman Lectures on Physics. Addison Wesley, Boston, Massachusetts (2005)
4. Fishburn, P. C.: Utility Theory for Decision Making. John Wiley & Sons Inc., New York (1969)
5. Fishburn, P. C.: Nonlinear Preference and Utility Theory. The John Hopkins Press, Baltimore, Maryland (1988)
6. Jaulin, L., Kiefer, M., Didrit, O., Walter, E.: Applied Interval Analysis, with Examples in Parameter and State Estimation, Robust Control, and Robotics. Springer, London (2001)
7. Jaynes, E. T., Bretthorst, G. L.: Probability Theory: The Logic of Science. Cambridge University Press, Cambridge, UK (2003)
8. Klir, G., Yuan, B. Fuzzy Sets and Fuzzy Logic. Prentice Hall, Upper Saddle River, New Jersey (1995)
9. Kreinovich, V.: Decision making under interval uncertainty (and beyond), In: Guo, P., Pedrycz, W. (eds.), Human-Centric Decision-Making Models for Social Sciences, Springer Verlag, pp. 163–193 (2014)
10. Luce, R. D., Raiffa, R.: Games and Decisions: Introduction and Critical Survey. Dover, New York (1989)
11. Mayer, G.: Interval Analysis and Automatic Result Verification. de Gruyter, Berlin (2017)

12. Mendel, J. M.: Uncertain Rule-Based Fuzzy Systems: Introduction and New Directions. Springer, Cham, Switzerland (2017)
13. Moore, R. E., Kearfott, R. B., Cloud, M. J. Introduction to Interval Analysis. SIAM, Philadelphia (2009)
14. Nguyen, H. T., Kosheleva, O., Kreinovich, V.: Decision making beyond Arrow's 'impossibility theorem', with the analysis of effects of collusion and mutual attraction. International Journal of Intelligent Systems **24**(1), 27–47 (2009)
15. Nguyen, H. T., Kreinovich, V., Wu, B., Xiang, G.: Computing Statistics under Interval and Fuzzy Uncertainty: Springer Verlag, Berlin, Heidelberg (2012)
16. Nguyen, H. T., Walker, C. L., Walker, E. A.: A First Course in Fuzzy Logic. Chapman and Hall/CRC, Boca Raton, Florida (2019)
17. Nilsson, H., Rieskamp, J., Jenny, M. A.: Exploring the overestimation of conjunctive probabilities. Fronters in Psychology **4**, Paper 101 (2013)
18. Novak, V., Perfilieva, I., Mockor, I.: Mathematical Principles of Fuzzy Logic. Kluwer, Boston, Dordrecht (1999)
19. Raiffa, H.: Decision Analysis. McGraw-Hill, Columbus, Ohio (1997)
20. Sheskin, D. J.: Handbook of Parametric and Non-Parametric Statistical Procedures. Chapman & Hall/CRC, London, UK (2011)
21. Thorne, K. S., Blandford, R. D.: Modern Classical Physics: Optics, Fluids, Plasmas, Elasticity, Relativity, and Statistical Physics. Princeton University Press, Princeton, New Jersey (2017)
22. Tversky, A., Kahneman, D.: Judgment under uncertainty: heuristics and biases. Science, New Series **185**(4157) 1124–1131 (1974)
23. Wedell, D. H., Moro, R.: Testing boundary conditions for the conjunction fallacy: effects of response mode, conceptual focus and problem type. Cognition **107**, 105–136 (2008)
24. Zadeh, L. A.: Fuzzy sets, Information and Control **8**, 338–353 (1965)