How to Combine Expert Estimates? How to Estimate Probability in the Intersection of Two Populations?

Miroslav Svítek[1], Olga Kosheleva[2],
Vladik Kreinovich[2], and Nguyen Hoang Phuong[3]
[1]Czech Technical University in Prague, Czech Republic, svitek@fd.cvut.cz
[2]University of Texas at El Paso, El Paso, Texas 79968
lbokati@miners.utep.edu, olgak@utep.edu, vladik@utep.edu
[3]Division Informatics, Math-Informatics Faculty, Thang Long University
Nghiem Xuan Yem Road, Hoang Mai District
Hanoi, Vietnam, nhphuong2008@gmail.com

1. The first problem

- Suppose that the experts E_1 and E_2 provided two estimates p_1 and p_2 for the probability of some event E.
- We would like to provide a single estimate that takes both estimates into account.
- To properly combine the two estimates, it is important to take into account how related are the opinions of the two experts.
- If can happen that in all previous situations and in this situation, the experts gave almost identical opinions.
- This probably means that they use the same technique to provide their estimates.
- In this case, the opinion of the second expert does not add anything new to the opinion of the first expert.
- So, the combined probability will still be the same value p_1.

2. The first problem (cont-d)

- It may happen that in the previous situations, the experts' opinions were independent.
- This means that they use different data and different techniques to estimate the probability.
- In this case, e.g.:
  - if both experts believe that this event is possible,
  - then taking both opinions into account should increase this probability.

3. The first problem (cont-d)

- On the other hand, if the expert opinions are negatively correlated, then we do not know whom to believe.
- Then, we should not take either of the probabilities seriously.
- In this case, the combined probability should be close to the do-not-know 0.5 value.
- In this talk, we show how to come up with a reasonable numerical value of the combined probability.

4. Formulation of the second problem

- Suppose that we know the frequencies of a certain phenomenon in two different populations.
- E.g., we know the frequency of a certain disease in a 50-60 age group and the frequency of this disease in women.
- What is the reasonable estimate for the frequency of this disease in the intersection of these two populations.
- E.g., among women in the 50-60 age bracket?
- In this paper, we show that this problem is mathematically similar to the first one.
- Thus, all the methods for solving the first problem can be automatically applied to the second problem as well.

5. Case of the first problem

- Let E be equal to 1 if the event happens and 0 if it does not.
- Let e_1 be equal to 1 if the first expert is correct in a randomly selected situation, and 0 if the first expert is wrong.
- Similarly, let e_2 be equal to 1 if the second expert is correct in a randomly selected situation, and 0 if the second expert is wrong.
- We know conditional probabilities $p(E \mid e_1) = p_1$, $p(E \mid e_2) = p_2$.
- Based on the analysis of the previous expert opinions, we can estimate the probability $p(e_1)$.
- We can do this by counting how many times the first expert was right.
- Similarly, based on the analysis of the previous expert opinions, we can estimate the probability $p(e_2)$.

6. Case of the first problem (cont-d)

- We can also estimate the probability $p(e_1 \,\&\, e_2)$ that both experts were right.
- Based on all this information, we want to estimate the probability $p(E)$.

7. What if we have only one expert?

- In this case, we know the conditional probability $p_1 = p(E \mid e_1)$ and the probability $p(e_1)$.
- We want to estimate the probability $p(E)$.

8. Second problem

- In the second problem, we only consider folks who belong to one (or both) of the two populations.
- Let E be equal to 1 if a randomly selected element has the desired phenomenon.
- Let e_1 be 1 if a randomly selected element belongs to the first population.
- Let e_2 be 1 if a randomly selected element belongs to the first population.
- We know the frequencies $p(E \mid e_1) = p_1$ and $p(E \mid e_2) = p_2$ of the phenomenon in each population.
- We know:
  - the number of elements n_1 in the first population,
  - the number of elements n_2 in the second population, and
  - the number of elements n_{12} that belong to both populations.

9. Second problem (cont-d)

- In this case, the overall number of elements that belongs to both populations is equal to $n_1 + n_2 - n_{12}$.
- Thus, we can estimate the probabilities of e_1, e_2, and $e_1 \,\&\, e_2$ as follows:

$$p(e_1) = \frac{n_1}{n_1 + n_2 - n_{12}}; \;\; p(e_2) = \frac{n_2}{n_1 + n_2 - n_{12}}; \;\; p(e_1 \,\&\, e_2) = \frac{n_{12}}{n_1 + n_2 - n_{12}}.$$

- Based on all this information, we want to find the probability $p(E \,|\, e_1 \,\&\, e_2)$.

10. Let us use the Maximum Entropy approach

- Situations in which we only have partial information about probabilities are ubiquitous.
- In such situations, several different probability distributions are consistent with our knowledge.
- In such cases, it makes sense not to pretend that our uncertainty is low.
- Thus, we should select the distribution with the largest possible uncertainty.
- A natural measure of uncertainty of a probability distribution is:
  - the average number of binary ("yes"-"no") questions that we need to ask
  - to fully determine which statements are true and which are false.

11. Let us use the Maximum Entropy approach (cont-d)

- It is known that this number is equal to Shannon's entropy

$$S = -\sum P_i \cdot \log_2(P_i).$$

- Here P_i are the probabilities of different possible situations.
- Thus, we need to select the distribution with the largest possible entropy.
- Such a selection is known as the Maximum Entropy approach.

12. What this means for the first problem

- In the first problem, we have three basic statement E, e_1, and e_2.
- Each of these statements is either true or false.
- Thus, we have $2^3 = 8$ possible situations:

$$E \,\&\, e_1 \,\&e_2, \quad E \,\&\, e_1 \,\&\neg e_2, \quad E \,\&\, \neg e_1 \,\&e_2, \quad E \,\&\, \neg e_1 \,\&\neg e_2,$$
$$\neg E \,\&\, e_1 \,\&e_2, \quad \neg E \,\&\, e_1 \,\&\neg e_2, \quad \neg E \,\&\, \neg e_1 \,\&e_2, \quad \neg E \,\&\, \neg e_1 \,\&\neg e_2.$$

- Let us use the following notations for their probabilities:

$$p_{111} \stackrel{\text{def}}{=} E \,\&\, e_1 \,\&e_2, \quad p_{110} \stackrel{\text{def}}{=} E \,\&\, e_1 \,\&\neg e_2, \quad p_{101} \stackrel{\text{def}}{=} E \,\&\, \neg e_1 \,\&e_2,$$
$$p_{100} \stackrel{\text{def}}{=} E \,\&\, \neg e_1 \,\&\neg e_2, \quad p_{011} \stackrel{\text{def}}{=} \neg E \,\&\, e_1 \,\&e_2, \quad p_{010} \stackrel{\text{def}}{=} \neg E \,\&\, e_1 \,\&\neg e_2,$$
$$p_{001} \stackrel{\text{def}}{=} \neg E \,\&\, \neg e_1 \,\&e_2, \quad p_{000} \stackrel{\text{def}}{=} \neg E \,\&\, \neg e_1 \,\&\neg e_2.$$

- These eight probabilities must add up to 1:

$$p_{111} + p_{110} + p_{101} + p_{100} + p_{011} + p_{010} + p_{001} + p_{000} = 1.$$

- We know the probability $p(e_1)$.

13. What this means for the first problem (cont-d)

- Thus, the fact that we know the value $p_1 = p(E \mid e_1) = p(E \,\&\, e_1)/p(e_1)$ is equivalent to knowing the probability

$$p(E \,\&\, e_1) = p_1 \cdot p(e_1).$$

- In terms of the basic probabilities, the probability $p(E \,\&\, e_1)$ has the form $p(E \,\&\, e_1) = p(E \,\&\, e_1 \,\&\, e_2) + p(E \,\&\, e_1 \,\&\, \neg e_2) = p_{111} + p_{110}$.
- Thus, we have $p_{111} + p_{110} = p_1 \cdot p(e_1)$.
- Similarly, we know the probability $p(e_2)$.
- Thus, the fact that we know the value $p_2 = p(E \mid e_2) = p(E \,\&\, e_2)/p(e_2)$ is equivalent to knowing the probability

$$(E \,\&\, e_2) = p_2 \cdot p(e_2).$$

- In terms of the basic probabilities, the probability $p(E \,\&\, e_2)$ has the form $p(E \,\&\, e_2) = p(E \,\&\, e_1 \,\&\, e_2) + p(E \,\&\, \neg e_1 \,\&\, e_2) = p_{111} + p_{101}$.
- Thus, we have $p_{111} + p_{101} = p_2 \cdot p(e_2)$.

14. What this means for the first problem (cont-d)

- Information about the values $p(e_1)$, $p(e_2)$, and $p(e_1 \,\&\, e_2)$ takes the following form: $p_{111} + p_{110} + p_{011} + p_{010} = p(e_1)$;
$$p_{111} + p_{101} + p_{011} + p_{001} = p(e_2); \quad p_{111} + p_{011} = p(e_1 \,\&\, e_2).$$
- So, to find the values p_{ikj}, we need to maximize – under the above constraints – the entropy
$$S = -p_{111} \cdot \log_2(p_{111}) - p_{110} \cdot \log_2(p_{110}) - p_{101} \cdot \log_2(p_{101}) -$$
$$p_{100} \cdot \log_2(p_{100}) - p_{011} \cdot \log_2(p_{011}) - p_{010} \cdot \log_2(p_{010}) -$$
$$p_{001} \cdot \log_2(p_{001}) - p_{000} \cdot \log_2(p_{000}).$$
- Entropy is a convex function of the probabilities, and the constraints are linear in terms of these probabilities.
- Thus, we can use the feasible convex optimization algorithms to find the desired probabilities.
- Once we find all the probabilities p_{ijk}, we can compute the desired probability $p(E)$ as $p(E) = p_{111} + p_{110} + p_{101} + p_{100}$.

15. Comments

- We can similarly consider the case when we have more than two experts and the case when we only have one expert.
- In the situation when we have only one expert, we have four possible situations $E \,\&\, e_1$, $E \,\&\, \neg e_1$, $\neg E \,\&\, e_1$, $\neg E \,\&\, \neg e_1$.
- Let us denote the probabilities of these situations by

$$p_{11} = p(E \,\&\, e_1),\ \ p_{10} = p(E \,\&\, \neg e_1),\ \ p_{01} = p(\neg E \,\&\, e_1),\ \ p_{00} = p(\neg E \,\&\, \neg e_1).$$

- These probabilities must add up to 1: $p_{11} + p_{10} + p_{01} + p_{00} = 1$.
- The available information – $p(e_1)$ and p_1 – lead to the constraints $p_{11} + p_{01} = p(e_1)$ and ${}_{11} = p_1 \cdot p(e_1)$.
- We can thus determine the probability p_{01} as

$$p_{01} = p(e_1) - p_1 \cdot p(e_1) = p(e_1) \cdot (1 - p_1).$$

- The only constraint on the remaining two values p_{00} and p_{10} is that all probabilities should add up to 1, so:

$$p_{00} + p_{10} = 1 - (p_{01} + p_{11}) = 1 - p(e_1).$$

16. Comments (cont-d)

- In this case, the maximum entropy approach leads to equal values of these two probabilities: $p_{00} = p_{10} = \dfrac{1 - p(e_1)}{2}$.
- Thus, the resulting estimate for the desired probability $p(E) = p_{11} + p_{10}$ has the form

$$p(E) = p_1 \cdot p(e_1) + \frac{1 - p(e_1)}{2} = \frac{1}{2} + p(e_1) \cdot \left(p_1 - \frac{1}{2}\right).$$

- This formula can be alternatively reformulated as

$$p(E) - \frac{1}{2} = p(e_1) \cdot \left(p_1 - \frac{1}{2}\right).$$

- In other words:
  - we should not take the expert estimate p_1 at face value,
  - we should adjust this estimate based on the expert's track record.

17. What this means for the second problem

- From the mathematical viewpoint, the two problems has similar inputs.
- The only two differences are as follows:
  - first, we only consider objects that belong to one of the populations, so
  $$p_{000} = p_{100} = 0;$$
  - second, what we want to estimate is different: $p(E \mid e_1 \,\&\, e_2)$.
- Thus, to solve the second problem, we:
  - perform the same optimization as in the first problem – with the additional constraint $p_{000} = p_{100} = 0$;
  - thus, we find the probabilities p_{ijk}, and then estimate:
  $$p(E \mid e_1 \,\&\, e_2) = \frac{p(E \,\&\, e_1 \,\&\, e_2)}{p(e_1 \,\&\, e_2)} = \frac{p_{111}}{p_{011} + p_{111}}.$$

18. Acknowledgments

- 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).
- It was also supported 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.