Conditional probabilities and Bayesian theorem in the study of animal disease introduction. (a didactic approach in environmental risk analysis) 

Daniela Laura Picardi

Department of Animal Production, Faculty of Agronomy, University of Buenos Aires. Av. San Martín 4453, Buenos Aires, Argentina

ABSTRACT

Bayesian analysis is one of the topics included in the subject “Environmental Risk Analysis” taught at the Faculty of Agronomy of the University of Buenos Aires, Argentina. Traditionally, we use a tree of probabilities to organize basic concepts necessary to understand this analysis (for example, conditional probability, excluding events, independence, etc.). Then, we deal with concepts of Bayesian analysis (such as additional conditional probability, prior probabilities, posterior probabilities, and their applications). In this work, we propose to analyze the problem method that starts by analyzing the scenarios.

• For that, we begin the analysis by paying attention to those events that we are certain have occurred.
• We will separate them according to their possible origins and assign them a probability (conditional).
• We will then use the probabilities to weight them, and then find a conditional probability of occurrence.

© 2022 The Author(s). Published by Elsevier B.V. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/)

ARTICLE INFO

Method name: Focus on scenarios
Keywords: Additional conditional probabilities, prior probabilities, posterior probabilities
Article history: Available online 29 September 2022

Specifications table

| Subject area: | Environmental Science |
| More specific subject area: | Environmental risk analysis |
| Name of your method: | Focus on scenarios |
| Name and reference of original method: | Study guide - Unit 4: “Bayesian Analysis”. Environmental Risk Analysis. Agronomy faculty. University of Buenos Aires. 2019. |
| Resource availability: | |

https://www.agro.uba.ar/
E-mail address: dpicardi@agro.uba.ar

https://doi.org/10.1016/j.mex.2022.101870
2215-0161/© 2022 The Author(s). Published by Elsevier B.V. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/)
The traditional method

About bayesian analysis

When we use Bayesian analysis in lessons, the main idea is to understand the conceptual framework for evaluating probabilities. It is traditionally explained (as part of this university course) that the probability of an event must be recalculated knowing that a new event happens and that both are not independent. We also explain that probabilistic resources are used to update our initial probabilistic assignment as we have new observations [3]. Thus, we seek to understand what circumstances determined that the data were those and not others through Bayesian induction.

Let’s look at the following definitions [2,3]:

- $P(B_i) = \text{prior}$
- $P(A_i/B_i) = \text{additional conditional probability}$
- $P(B_i/A_i) = \text{posterior}$

The simplest way to express the Bayes rule is:

$$P(B/A) = \frac{P(A/B)P(B)}{P(A)}$$  \hspace{1cm} (1)

And since

$$P(A) = \sum_{i=1}^{k} P(A/B_i)P(B_i)$$  \hspace{1cm} (2)

Then

$$P(B/A) = \frac{P(A/B)P(B)}{\sum_{i=1}^{k} P(A/B_i)P(B_i)}$$  \hspace{1cm} (3)

Tree methodology

Let us now consider a scenario that frequently occurs in risk analysis, the importation of animals. Usually, we make a tree diagram to organize the calculations linked to the probabilities involved. For example:

The probability of reaching the end of any complete path is the product of the probabilities in its branches. The prevalence of a disease constitutes the prior probability (simple probability) that a proportion of animals are infected with a particular disease. The additional information (conditional probability) is constituted by the probabilities of the quality of the test that detects the disease...
(sensitivity, specificity, false positives, and false negatives) [3]. Finally, we can calculate, for example, the probability that an animal is infected, since a specific veterinary control is negative (a posterior probability).

Focus on scenarios (a change to the traditional method)

As we have seen, in environmental risk analysis we need to answer questions using probabilistic values. Traditionally one of the proposed methodological approaches is to answer these questions through Bayesian analysis, and in particular the use of the probability tree seen above [1]. First, the prior probabilities are identified and located. Second, the additional conditional probability, and third, the posterior probabilities are computed. Then (fourth step) these probabilities are in the tree and finally (fifth step) the Bayes theorem is applied using the appropriate probabilities for the case.

In this case, we propose a method based on obtaining these answers (probabilities) through a similar analysis (because it uses the Bayes theorem) but focused on starting by thinking about the situations that originate the observed event. That is, as a first step, look at the nature of the events and then determine what are the appropriate conditional probabilities for each situation.

Going back to the example case (the importation of animals), we apply this method of approach by scenarios to quantify probabilities and show the proposed reasoning. We needed to know the probability that an animal is infected, since a specific veterinary control is negative (a posterior probability).

Method steps

1. Define the question of interest and think about the conditionality of the events:
   “¿What is the probability that an animal is infected, since a specific veterinary control is negative?”
   • This question is associated with a conditional probability: \( P(B/A) \)
   • In this case: \( P(\text{infected animal/negative test}) \).

   √ “Infected animal” is one of the possible states of the “animal health” situation.

   And it depends on:

   √ “Negative test” is one of the possible states of the “test result” situation.

2. In the previous question, ¿what has surely happened?
   • The test result is negative.
   • This information is indicated after the “/” symbol. In this case “/negative test”.

3. Define possible events that cause the one defined in “2” (in the example, negative result).

   | Event 1 | Event 2 |
   |---------|---------|
   | The test is negative AND the animal tested is healthy. there is no evidence of the presence or absence of the disease. | The test is negative AND the animal tested is infected but is not being detected. |
   | This situation is desirable. It is related to the quality of the test (Specificity) | It is related to the quality of the test. (F-) |
   | Specificity → P (negative test/healthy animal) | F- → P (negative test/infected animal) |
   | Simple Probability: P (healthy animal) | Simple Probability: P (infected animal) |

   We have defined two situations that originate what is indicated after “/”. (That is, two situations associated with a negative test): we are faced with an error (F-) or the test correctly marked the absence of the disease.

4. Let us now recall the question of interest: “¿What is the probability that an animal is infected, since a specific veterinary control is negative?”. To answer, let us quantify the probability of occurrence of a particular situation in relation to all the situations that could cause a negative result of the test (in our example those are the situations defined in “3”). This situation defined in the numerator must be the one that answers the question of interest.
We are facing an event 2
\[ \frac{ \text{We are facing an event 2} }{ \text{We are facing an event 1} } = \frac{F2}{F2 + F1} \]  \hspace{1cm} (4)

5. Finally, after starting our analysis thinking about events and their conditionality, we will use probabilities to solve equations.

So, in (4), for each of these events \((F_i)\) we must consider a \textbf{joint probability}. Recall that when two events are not mutually exclusive and are not independent (that is, there is conditionality), the joint probability is obtained as \([2]\):

\[ P(A \text{ and } B) = P(A/B) * P(B) \]  \hspace{1cm} (5)

Now we will write the terms of (4) following the logic of (5)

\[ F1 = P(\text{negative test AND healthy animal}) = P(\text{negative test/healthy animal}) * P(\text{healthy animal}) \]  \hspace{1cm} (6)

\[ F2 = P(\text{negative test AND infected animal}) \]
\[ = P(\text{negative test/infected animal}) * P(\text{infected animal}) \]  \hspace{1cm} (7)

To simplify the notation, we will rename the events:

| Events         | Possible states                      |
|----------------|--------------------------------------|
| A Test result  | A1 (Negative test) A2 (Positive test) |
| B Animal health| B1 (Infected animal) B2 (Healthy animal) |

Finally, using (6) and (7) in (4) we obtain that

\[ P(\text{infected animal/negative test}) = P(B_1/A_1) = \frac{P(A_1/B_1)P(B_1)}{[P(A_1/B_2)P(B_2)] + [P(A_1/B_1)P(B_1)]} \]  \hspace{1cm} (8)

Let’s express the denominator as a sum where “i” indicates the events that could cause a negative test \((A_i)\):

\[ P(\text{infected animal/negative test}) = \frac{P(A_1/B_1)P(B_1)}{\sum_{i=1}^{2} P(A_1/B_i)P(B_i)} \]  \hspace{1cm} (9)

\textbf{Conclusion}

We have applied the method proposed in steps “1” to “4” after analysing the nature of the events. We can see that by answering the question of interest by quantifying the events as in (4) we are returning to Bayes’ theorem. In this way we can validate that the result obtained with this method is the same as the traditional method. Both end up getting an answer through a Bayesian analysis.

In other words, we do not get the answer to the question of interest by starting by setting up the numerical values of probabilities in a tree diagram and then thinking about which ones to use (as we solve it traditionally). Instead, we begin by analysing the question we are trying to answer by first understanding the logic of the related events. After this analysis, we think about what type of probability is appropriate to that situation and what is the causality. Finally, we organize the information and only in “5” do we give numerical values to the probabilities to obtain the answer we are looking for.

This is a didactic quantitative method to quantify and understand the appropriate use of statistical probabilities and the usefulness of the Bayes theorem. This focuses on the process on understanding the events first and then on how the Bayes theorem solves them, instead of starting by explaining the theorem and what probabilities make it up.
Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

The author gratefully acknowledges the assistance of the Dra. Lic. Eliana Belen Gonzalez during the writing process. I am also grateful to Mg. Sc. Engineer Laura Pruzzo for contributing her knowledge. Both teachers have been colleagues in teaching the subject Environmental Risk Analysis.

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

References

[1] M. Daniel Kammen, M. David Hassenzahl, Should We Risk It? Exploring Environmental, Health, and Technological Problem Solving, Princeton University Press, 2001, doi:10.1515/9780691188317.
[2] L. Mesa, M. Lozano y, J. Romero, Descripción general de la Inferencia Bayesiana y sus aplicaciones en los procesos de gestión, La simulación al servicio de la academia 2 (2011) 1–28 https://www.urosario.edu.co/Administracion/documentos/investigacion/laboratorio/miller_2_2.pdf.
[3] Study guide, Environmental Risk Analysis. Agronomy faculty, University of Buenos Aires, 2019.