

Simplifying Bayesian Inference

Stefan Krauß, Laura Martignon & Ulrich Hoffrage
Max Planck Institute For Human Development
Lentzeallee 94, 14195 Berlin-Dahlem

Probability theory can be used to model inference under uncertainty. The particular way in which Bayes' formula is stated, which is of only minor importance in standard probability textbooks, becomes central in this context. When events can be interpreted as evidences and hypotheses, Bayes' formula allows one to update one's belief in a hypothesis in light of new data.

Is unaided human reasoning Bayesian?

Kahneman and Tversky (1972) affirmed: "In his evaluation of evidence, man is not Bayesian at all." In their book *Judgment under uncertainty* (1982), they attempted to prove that human judgment is riddled with systematic deviations from the logical and probabilistic norm. In chapter 18 of the same book David M. Eddy stressed that medical doctors do not follow Bayes' formula when solving the following task:

The probability that a woman at age 40 has breast cancer (B) is 1%. ($P(B) = \text{prevalence} = 1\%$)

According to the literature, the probability that the disease is detected by a mammography (M) is 80%. ($P(M+ | B) = \text{sensitivity} = 80\%$)

The probability that the test misdetects the disease although the patient does not have it is 9.6%. ($P(M+ | \neg B) = 1 - \text{specificity} = 9.6\%$)

If a woman at age 40 is tested as positive, what is the probability that she indeed has breast cancer ($P(B|M+)$)?

Bayes' formula yields the following result:

$$P(B | M+) = \frac{P(M+ | B) \cdot P(B)}{P(M+ | B) \cdot P(B) + P(M+ | \neg B) \cdot P(\neg B)} = \frac{80\% \cdot 1\%}{80\% \cdot 1\% + 9.6\% \cdot 99\%} = 0.078$$

Thus, the probability of breast cancer is only 7.8%, while Eddy reports that 95 out of 100 doctors estimated this probability to be between 70% and 80%.

Gigerenzer and Hoffrage (1995) focused on another aspect of the problem: the representation of uncertainty. In Eddy's task, quantitative information was given in probabilities. Gigerenzer and Hoffrage presented Eddy's problem to medical doctors replacing probabilities with a different representation of uncertainty, namely *natural frequencies*.

In their formulation the task was:

100 out of every 10000 women at age 40 who participate in routine screening have breast cancer.

80 of every 100 women with breast cancer will get a positive mammography.

950 out of every 9900 women without breast cancer will also get a positive mammography.

Here is a new representative sample of women at age forty who get a positive mammography in routine screening. How many of these women do you expect to actually have breast cancer?

Now nearly half (46%) of all doctors gave the Bayesian answer: 80 out of 1030 (7.8%).

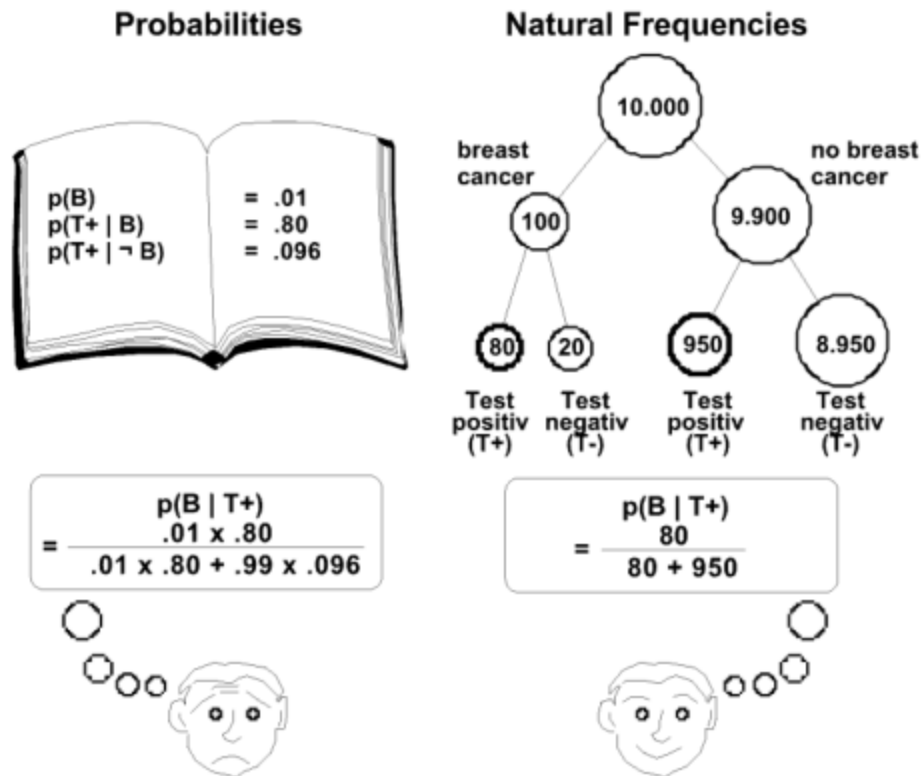


Figure 1

What is the crucial property that helps one to find the Bayesian solution? To answer this question, it is helpful to consider a more general case. In real-life situations, decisions are usually based on several cues. A medical doctor, for instance, seldom diagnoses a disease based on a single test. The usual procedure after a mammography is to perform an ultrasound test (U). For an ultrasound test, sensitivity and specificity are usually given in the instructions:

$$P(U+ | B) = 95\%$$

$$P(U+ | \neg B) = 4\%$$

In an empirical study, we presented this information together with $P(B)$, $P(M+ | B)$ and $P(M+ | \neg B)$ to a group of participants. They were asked: What is the probability that a woman at age 40 has breast cancer, given that she has a positive mammography *and* a positive ultrasound test?

When given this probability format, only 12.2% of our participants reached the correct solution ($\gg \frac{2}{3}$).

D. Massaro (1998) gave an example describing the same situation with frequencies¹:

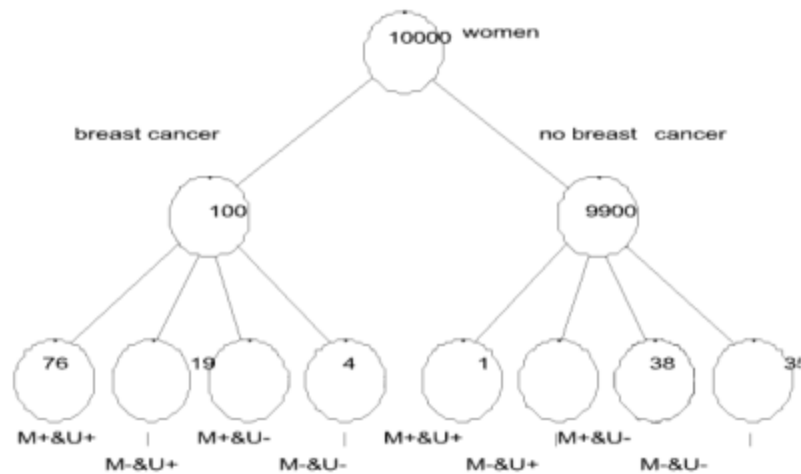


Figure 2

Massaro writes that in the case of two cues „a frequency algorithm will not work” and „it might not be reasonable to assume that people can maintain exemplars of all possible symptom configurations.”

However, his statements are not based on experimental evidence, and his frequency configuration is not really equivalent to the probability format because he works with combined sensitivity $P(M+ \& U+ | B)$ and combined specificity $1 - P(M- \& U- | \neg B)$.

One possible frequency format, which does correspond to our probability format, is²:

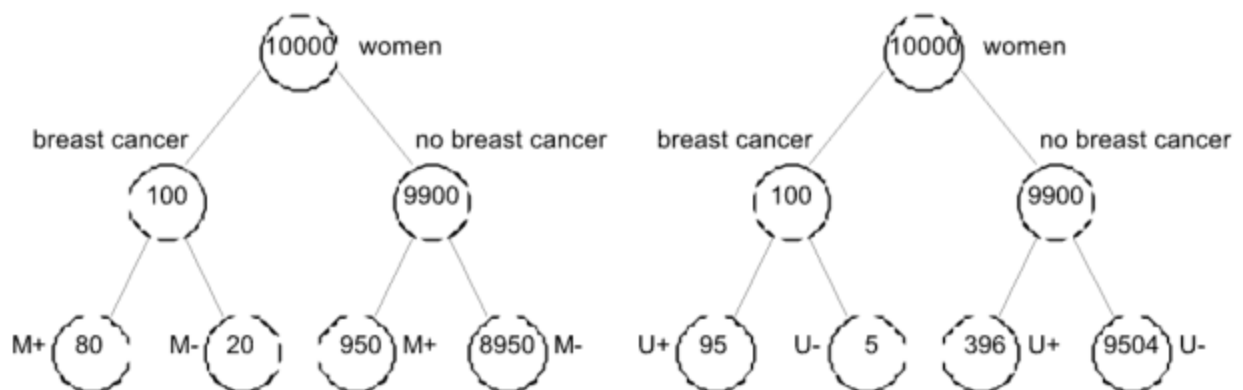


Figure 3

In words:

100 out of every 10000 woman at age 40 who participate in routine screening have breast cancer.

80 of every 100 women with breast cancer will get a positive mammography.

950 out of every 9900 woman without breast cancer will also get a positive mammography.

95 out of 100 women with cancer will get a positive ultrasound test.

396 out of 9900 women, although they do not have cancer, nevertheless obtain a positive ultrasound test.

How many of the women who get a positive mammography and a positive ultrasound test do you expect to actually have breast cancer?

14.6% of our participants solved this version correctly.

Another possibility is to consider the tests sequentially. This is possible because the ultrasound test and the mammography are conditionally independent, i.e. $P(U+ | B) = P(U+ | B \& M+)$. Now we have³:

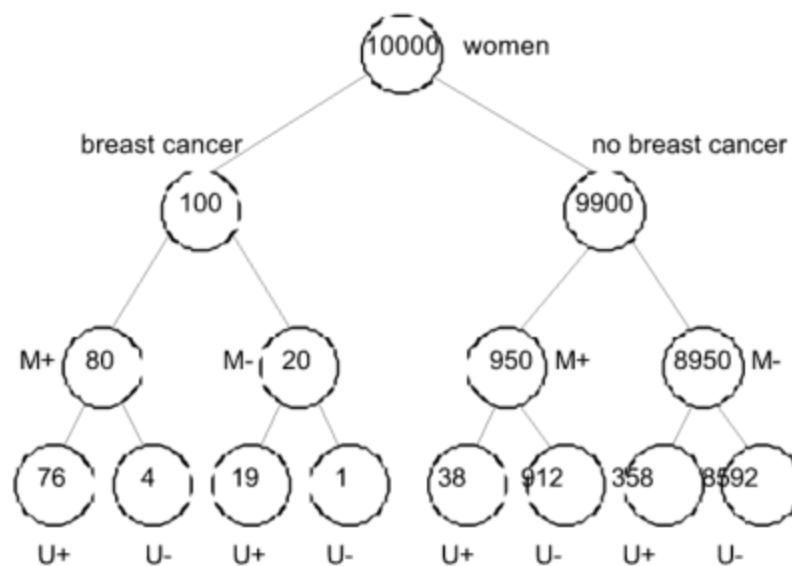


Figure 4

In words:

100 out of every 10000 women at age 40 who participate in routine screening have breast cancer.

80 of every 100 women with breast cancer will get a positive mammography.

950 out of every 9900 women without breast cancer will also get a positive mammography.

76 out of 80 women who had a positive mammography and have cancer also have a positive ultrasound test.

38 out of 950 women who had a positive mammography, although they do not have cancer, also have a positive ultrasound test.

How many of the women who get a positive mammography and a positive ultrasound test do you expect to actually have breast cancer?

53.7% of our participants solved this task correctly.

Not all frequencies in the tree were actually used. The next step is to eliminate all frequencies irrelevant to the task. Thus we obtain:

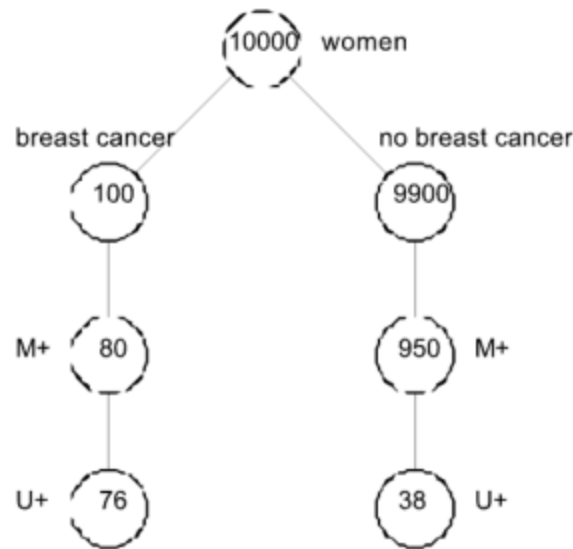


Figure 5

These frequencies, namely those that really foster insight, deserve a special name. We decided to call them *Markov frequencies* because of the natural analogy with Markov chains. In fact:

- 1) Our tree consists of two *chains* which are joined at the root.
- 2) Each node corresponds to the reference class that determines the next node. As in a Markov chain, the frequency in each node depends only upon its predecessor, not upon previous nodes.

Being able to “think in chains” seems crucial for human insight and fits the modern view that problem solving, unlike perception, is sequential rather than parallel. Markov frequencies are task-oriented, i.e., only information that is relevant for the task appears in the tree. Gigerenzer and Hoffrage (1995) also used a tree (see Figure 1). Their tree contains the information $(P(T-|B) \text{ and } P(T-|\neg B))$, which is not relevant to the question “ $P(B|T+) = ?$ ”. In our chains, the odds of the problem can be read directly from the last two nodes. This is because the tree with Markov frequencies corresponds to the well-known likelihood-combination rule (see, for instance, Spies, 1993):

$\text{prior odds} \cdot \text{product of the likelihood ratios} = \text{posterior odds}$

The prior odds for breast cancer are $\frac{100}{9900}$.

Multiplying this with the likelihood ratio for the mammography⁴, we obtain $\frac{80}{950}$.

Again multiplying this with the likelihood ratio of the ultrasound test, we finally get $\frac{76}{38}$.

By using Markov frequencies, it is not only clear which information should be given to experts, but also which information should be omitted⁵. Appropriately deleting useless information is part of the overall computation, as we know from information theory.

References

- Eddy, D. M. (1982). Probabilistic reasoning in clinical medicine: Problems and opportunities. In D. Kahneman, P. Slovic & A. Tversky (Eds.), *Judgment under uncertainty: Heuristics and biases* (pp. 249-267). Cambridge, England: Cambridge University Press.
- Gigerenzer, G. & Hoffrage, U. (1995). How to improve bayesian reasoning without instruction: Frequency formats. *Psychological Review*, 102, 684-704.
- Kahneman, D. & Tversky, A. (1972). Subjective probability: A judgement of representativeness. *Cognitive Psychology*, 3, 430-454.
- Massaro, D. (1998). *Perceiving talking faces* (pp.174-179). Boston. MIT Press.
- Spies, M. (1993). *Unsicheres Wissen: Wahrscheinlichkeit, Fuzzy-Logik, neuronale Netze und menschliches Denken* (pp.51-54). Heidelberg, Berlin, Oxford: Spektrum Akademischer Verlag.

Footnotes

- 1) To integrate the research on this topic, we borrowed concepts from various sources and explored them in the breast cancer example. In fact, Gigerenzer and Hoffrage used a sample of 1.000 (not 10.000) women, Massaro speaks of symptoms instead of tests and we tested our subjects with „tuberculosis tasks“ instead of „breast cancer tasks.“
- 2) Gigerenzer and Hoffrage stressed that only frequencies work that can be sampled „naturally“. A doctor would get information of this kind when he samples *instructions* for different tests and translates the information therein into frequencies.
- 3) A doctor would get information of this kind when he samples patients with respect to their state of illness.
- 4) The likelihood ratio $L(B, M+)$ is defined by $\frac{P(M+ / B)}{P(M+ / -B)}$, which is $\frac{80\%}{9,6\%} \gg 8.3$

The likelihood ratio $L(B, U+)$ therefore is $\frac{95\%}{4\%} = 23.75$
- 5) Because Bayes' formula can be used to model inference under uncertainty, it is also a tool in scientific reasoning. Klaus Hasselmann from the Max Planck Institute for Meteorology in Hamburg is presently applying a Bayesian analysis to hypotheses about changes in climate. The Society for Mathematics and Data Analysis in St. Augustin is investigating various methods for

estimating credit risks, such as analysis of discriminance, fuzzy-pattern classification, and neural networks with the help of Bayes' theorem. The „Krebsatlas“ (almanac of cancer patients) for Germany is being reviewed at the Ludwig Maximilian University in Munich by means of Bayesian methods. The task is to detect and eliminate spurious correlations. Even the Microsoft Office Assistant uses Bayesian procedures. The mathematician Anthony O' Hagan „elicits“ on behalf of the British government hydrological conductivity of the rock at Sellafield from experts. He uses their beliefs to determine a prior distribution, with which the appropriateness of the area as a permanent disposal site for nuclear waste can be estimated (Neue Zürcher Zeitung, May 13, 1998, S.39.). Even the most expert systems are based on Bayes' formula. A famous example is MUNIN (Muscle and Nerve Inference Network) from Lauritzen and Spiegelhalter (1988), which is used for making diagnoses on the basis of measurements of muscular electrical impulses („electromyography“). Maybe Markov frequencies can also help to facilitate programming those expert systems.

Acknowledgments

We thank Valerie Chase, Martin Lages, Donna Alexander and Matthias Licha for helpful comments and Ursula Dohme for running the experiments.

(Submission to the 1998 Conference on „Model-Based Reasoning in Scientific Discovery“)