

Concept paper

Assessing Scientific Evidence of Singular Historic Events by Bayes Theorem

Christian Damgaard

Department of Ecoscience, Aarhus University

Abstract: The typically qualitative treatment of evidence underlying different causal hypotheses that attempt to explain singular historic events is criticized. Instead of using *ad hoc* qualitative arguments, it is here advocated to use Bayes theorem to assess how well evidence either support or weaken a specific causal hypothesis. It is argued that such a practice will enable scientists who study singular historic events to discuss, compare and aggregate different beliefs in a meaningful way.

Keywords: Bayes theorem; singular historic events

Introduction

In the scientific exploration of a phenomenon, we often encounter a situation where different hypotheses may explain the phenomenon. To address which of the possible hypotheses that are most likely to be true, we typically adopt a research strategy of compiling evidence that either support or weaken a specific hypothesis. In many scientific investigations, it is possible to sample from the studied phenomenon and get independent but comparable pieces of information that may be treated mathematically using a sampling theoretic approach, and the different hypotheses may be compared using statistical methods. However, often the study object is a unique singular historic event, e.g. the origin of life, the extinction of the mammoth or the Yamnaya invasion of Europe, where we want to understand the causal mechanisms that lead to the event, and it is by the singular nature of the study object not possible to sample from the phenomenon.

When studying singular historical events, pieces of evidence are collected and the possible hypotheses are assessed in the light of the compiled evidence. The compilation of evidence is often a costly enterprise, but unfortunately the large effort that is put into collecting evidence is typically not matched by a similarly rigorous assessment of the evidence. For example, different hypotheses have been suggested to explain the relatively sudden and large social and economic changes in the European bronze age (Anthony 2010) accompanied by large-scale population migrations and replacements (Allentoft et al. 2015), which is known as the Yamnaya invasion of Europe. One of the hypotheses is that the invasion was facilitated by the occurrence of the plague (Rasmussen et al. 2015), and based on discoveries of the plague bacteria in ancient DNA the authors claim that this hypothesis is indeed plausible. What do the authors mean by “plausible”, and how is this hypothesis compared to the other existing hypotheses that attempt to explain the Yamnaya invasion?

The above-mentioned example of a causal hypothesis that is described as being “plausible” is typical. Generally, the collected evidence underlying different causal hypotheses that attempt to explain singular historic events are assessed by *ad hoc* qualitative arguments that result in statements like “event A is most likely caused by B”. This practice is problematic, since *ad hoc* methods of assessing evidence face the risk of incorporating preconceived judgment into the assessment process in a nontransparent manner. Furthermore, the use of *ad hoc* methods to assess the collected evidence is not the best possible practice, since a general, logically consistent and quantitative method for assessing evidence exists, i.e. by applying Bayes theorem (Jaynes 2003).

Bayes theorem used on singular historic events

Bayes theorem specifies that,

$$P(H|E) = \frac{P(H) P(E|H)}{P(E)} \quad (1),$$

where H is a hypothesis and E is evidence. The probability of the hypothesis being true after the evidence has been considered, i.e. the posterior probability $P(H|E)$, is a function of the prior probability of the hypothesis being true, $P(H)$, the probability of observing the evidence assuming the hypothesis is true, $P(E|H)$ and the marginal probability of observing the evidence, $P(E)$.

The prior probability of the hypothesis being true is a measure of our belief in the hypothesis without considering the new evidence. This measure of belief is subjective and depends on the investigator, but the advantage is that if the belief is assigned a probability, then the subjective opinion becomes transparent and it becomes possible to discuss, compare and aggregate different beliefs in a meaningful way. If there are more pieces of evidence that are relevant for a hypothesis, then Bayes theorem may be used multiple times, where the posterior probability, $P(H|E)$, is used as the updated prior distribution of the hypothesis, $P(H)$, when assessing the next piece of evidence (Jaynes 2003).

The probability of observing the evidence assuming the hypothesis is true, $P(E|H)$, which is equivalent to the likelihood function in statistical analyses, is the key step in the quantitative assessment of the evidence. A typical problem when making qualitative *ad hoc* assessments of evidence is to make the implicit assumption that $P(E|H) \gg P(E)$, so that the observation of a specific piece of evidence is interpreted as an almost proof that the hypothesis is true, i.e. $P(H|E) \approx 1$. However, if this implicit assumption is incorrect, then the posterior probability of the hypothesis is, to some degree, controlled by the typically unspecified prior probability of the hypothesis. In the limiting pathological case, when observing the evidence does not depend on the hypothesis and $P(E|H) = P(E)$, the posterior probability is equal to the prior probability, $P(H|E) = P(H)$.

More generally, the ratio $P(E|H)/P(E)$ determines the amount by which the prior hypothesis is altered by observing E , and it is often simpler to quantify the odds ratio between the likelihood of observing a piece of evidence under two complementary hypotheses. The odds ratio of the posterior probabilities of hypothesis A and the complementary hypothesis \bar{A} (not A) is,

$$\frac{P(H_A|E)}{P(H_{\bar{A}}|E)} = \frac{P(H_A)}{P(H_{\bar{A}})} \frac{P(E|H_A)}{P(E|H_{\bar{A}})} \quad (2),$$

where the marginal $P(E)$ term cancels out. Consequently, if we can determine how much more likely it is to observe evidence E under hypothesis A compared to under the complementary hypothesis \bar{A} , then we may calculate the posterior probability for hypothesis A for a given prior probability of the hypothesis.

The outlined quantitative procedure for assessing the evidence for different hypotheses of the causal explanation that determine singular historical events is by no means an original suggestion. Already in the 14th century, the monk Nicolas Autrecourt proposed to compare different hypotheses by their probability of being correct (Ebbesen 2006), and the general use of Bayes theorem as a quantitative extension of Aristotelian logic has most convincingly been advocated by Jaynes (2003).

Case studies

Simple conceptual example

Assume that we investigate a phenomenon where there are only two mutually exclusive hypotheses (H_A and $H_B = H_{\bar{A}}$) that explain the phenomenon. From the outset of the investigation we have no prior preference for either hypothesis, i.e., $P(H_A) = P(H_{\bar{A}}) = 0.5$. Then we observe a piece of evidence that increases our belief in hypothesis A , and the ratio $P(E|H_A)/P(E)$ is assessed to 1.5, then $P(H_A|E) = 0.75$.

Alternatively, using the ratios of the posterior probability distributions (eq. 2), we have that $\frac{P(H_A)}{P(H_{\bar{A}})} = 1$, and $\frac{P(E|H_A)}{P(E|H_{\bar{A}})} = \frac{0.75}{0.25} = 3$. That is, the likelihood of observing evidence E is three times higher under hypothesis A compared to under the complementary hypothesis \bar{A} .

No evidence

Sometimes it is argued that the absence of evidence supports a specific hypothesis. In such cases, it is important to investigate the probability of observing the evidence or, rather, not observing the evidence, $P(!E)$, where $P(!E) = 1 - P(E)$. If the probability of finding such evidence is close to zero, e.g. due to a poor fossil record, then the absence of evidence does not lead any support to the hypothesis. $P(E) \approx 0 \Rightarrow P(!E) \approx 1 \Rightarrow P(!E|H)/P(!E) \approx 1$

The Yamnaya invasion of Europe

As mentioned above, it has been speculated that the Yamnaya invasion of Europe was facilitated by the plague (Rasmussen et al. 2015). However, the pest bacteria could have spread to the resident European human population before the Yamnaya invasion, and the discovery of the pest bacteria within the Yamnaya population is necessary, but not sufficient, evidence for the hypothesis that the Yamnaya invasion of Europe was facilitated by the infection of the resident European human population with plague.

The odds ratio as well as the prior probability distribution for the plague hypothesis need to be assessed by experts, but it may be fair to assume that the odds ratio < 10 , and that the most likely prior hypothesis is that the invasion was made possible by the use of horses and a surprise attack strategy on relatively small rural settlements. If the prior probability of the hypothesis is 0.1 and the odds ratio is 5, then the posterior probability in favor of the plague hypothesis is $\frac{0.1}{0.9} \frac{5}{1} = 0.56$.

Moreover, if a body of experts meet to discuss the most likely values of the above-mentioned odds ratio and prior probability of the different hypothesis, then the uncertainty of the assessment may be quantified by the distribution of the posterior probabilities.

Discussion

In many archeological and historical studies, it is possible to investigate a general phenomenon by sampling independent events or sites. For example, Armit et al. (2014) deduced by quantitative methods that the societal collapse at several Bronze age sites was not caused by climate change. Therefore, the criticism brought forward here is not a general criticism of the archeological and historical sciences, but only studies of singular historic events where the rigorous and quantitative assessment of the compiled evidence is lacking.

Moreover, it is unsatisfactory that the disagreement and uncertainty within the scientific community regarding the status of the different hypotheses that explain singular historic events are only discussed and communicated to the broader public by qualitative arguments. The lack of quantitative arguments has led to a history of unnecessary exaggerations and trench digging, which is scientifically counterproductive and damaging for the reputation of science in the eyes of the public. Instead, it is argued here that the combined uncertainty of the scientific community is best evaluated and communicated by letting a body of experts meet to quantify the uncertainty of the assessment by distributions of the posterior probabilities of different hypotheses.

Acknowledgements: Thanks to Mette Løvschal for valuable suggestions in preparing this paper.

References

- Allentoft ME, Sikora M, Sjögren K-G, Rasmussen S, Rasmussen M, Stenderup J, Damgaard PB, Schroeder H, Ahlström T, Vinner L et al. 2015. Population genomics of bronze age eurasia. *Nature*. 522(7555):167-172.
- Anthony DW. 2010. *The horse, the wheel, and language*. Princeton University Press.
- Armit I, Swindles GT, Becker K, Plunkett G, Blaauw M. 2014. Rapid climate change did not cause population collapse at the end of the european bronze age. *Proc Natl Acad Sci*. 111(48):17045.
- Ebbesen S. 2006. Skolastikken. In: Jensen HS, Knudsen O, Stjerntfeldt F, editors. *Tankens magt*. København: Lindhardt & Ringhof. p. 385-410.
- Jaynes ET. 2003. *Probability theory: The logic and science*. Cambridge: Cambridge University Press.
- Rasmussen S, Allentoft Morten E, Nielsen K, Orlando L, Sikora M, Sjögren K-G, Pedersen Anders G, Schubert M, Van Dam A, Kapel Christian Moliin O et al. 2015. Early divergent strains of yersinia pestis in eurasia 5,000 years ago. *Cell*. 163(3):571-582.