Using Bayesian statistics in historical research

Heidi Suurkaulio

May 10, 2022

Bayesian statistics is not method I use in my masters' thesis. Instead, I got interested about it and its possibilities in historiography during the introduction course to statistics lectured by Harri Högmander. Högmander himself has participated in writing an article, I will later refer as an example of Bayesian Statistics utilized in historical census (demographic studies).

The other even more influential motive have been my discussions with my spouse. My spouse Joni Pääkkö currently is writing his masters' thesis about Bayesian approaches to psychoacoustics (study of e.g. human hearing thresholds). His further work can be checked here: `https://github.com/joanpaak`.

To make it clear, I am in no means a professional or advanced in Bayesian statistics. But I do have the basic conceptual understanding about the method. This essay does not include a comprehensive picture of the field or clear tutorial to the approaches. Instead, I try to explain the basic concept the best I can, and evaluate the pros and cons of Bayesian approaches to historical studies.

I will start my essay by introducing Bayesian statistics and the differences it has with frequentistic statistics. Then I briefly cover the history of Bayesian statistics. Lastly I focus on the problems and benefits of using Bayesian statistics in the historical research.

1 What is Bayesian statistics

Similarly to other statistical and numerical methods, Bayesian statistics is a quantitative method. Bayesian method shares some concepts, such as different distributions[1], mean and standard deviation, with other statistic approaches. However, the fundamental difference is the concept of probability. The key concept in Bayesian statistics is formula, which quantifies how the (prior) information is changed with the emergence of new data.

$$P(\theta|y) \propto P(\theta)P(y|\theta)$$ [2]

There are multiple variation of the formula,[3] but I personally like this one, because it is somewhat clear. With a natural language it means: "probability (P) of hypothesis (θ theta) given data (y) is proportional to the prior probability of hypothesis multiplied by probability of data given the hypothesis."

From now on, I am going to compare only the Bayesian and frequentistic (or classical) statistics and ignore other possible quantitative methods to be clear. The classical probability means that if we have a finite amount of events a possibility to one certain event is the amount of events divided by the amount of the certain one.[4] For example probability of getting value 3 when tossing a die is 1/6.

In Bayesian approach the possibility is determined "the other way around". The main question being, what has caused the event. As John K Kruschke put it, if we see a wet sidewalk, we may consider possible causes for that. Looking around we gather new evidence, for example rain clouds or irrigation device. If we have enough evidence we can rule out the other possible causes and stick with the most probable one. According to Kruschke, one analog to Bayesian inference could also be Sherlock Holmes ruling out possible committees of the crime and ending up in the last credible option.[5] Bayesian statistics is mathematical formalization of this inference.

In other words: we may think that the event has already happened, like some historical event. For example, we can think about the gender of viking warriors. In the turn of the 20th century it was widely believed that all the warriors were men. However, with new archaeological findings there is some evidence of women buried with weapons. So the

[1]The most famous distribution being normal distribution (or Gaussian), there are also other distributions such as gamma- and Poisson distribution.
[2]Gelman et al (2014) p 7.
[3]See for example: `https://www.mathsisfun.com/data/bayes-theorem.html` or Kruschke (2014) p 100 - 101.
[4]Matematiikkapakki: todennäköisyyslaskenta (9.5.2022).
[5]Kruschke (2014) p 16 - 19.

assumptions about the gender have shifted with the new evidence. Of course, it is not plausible that all of the warriors or even half of the warriors had been women. But, at least we think that some of them were. This is called *subjective* or *inverse* probability. As with history, with Bayesian statistics the uncertainty is taken into consideration.

Whereas in frequentistic statistics parameters are used with (fixed) values, in Bayesian approach the parameters are defined with distributions.[6] Other key differences are also that in Bayesian statistics statistical significance (or the p-value) is not in an outstanding position. Also the null hypotheses are not evaluated in the "familiar way".[7]

According to Kruschke Bayesian analysis basically follows these five steps:

1. Finding and identifying the suitable data for the research question.
2. Defining the descriptive behavior for the data (scale and distributions).
3. Specifying the prior distribution for the data.
4. Doing the actual statistical analysis which creates the posterior distribution (numeric approximation).
5. Checking if the posterior distribution really mimics the original data (with simulations).[8]

Bayesian statistics is older than one may think. According to statistician Stephen E. Fienberg (27.11.1942 – 14.12.2016)[9] the Bayesian statistics lies on the Bayes' theorem which is said to be written by a English minister (churchman) Thomas Bayes (c. 1701 - 1761)[10]. The paper *An Essay Towards Solving a Problem in the Doctrine of Chances* was published posthumously in 1763. Nevertheless, it is controversial whether or not the paper actually contains the "Bayes' theorem". It is also ambiguous was the paper even written by Bayes and could we call Bayes "Bayesian" in the modern sense.

A much more influential and elaborated version of the Bayes' theorem was written by a French scientist Pierre-Simon Laplace (23.3.1749 – 5.3.1827)[11] in 1774. Overall, the question of *inverse probability* (which I will discuss later) interested scientists throughout the 19th century and the beginning of 20th century. Yet, in the first half of the 20th century a more frequentistic approach of statistic became the "mainstream" by the

[6] Kruschke (2014) p 19 - 22.
[7] See for example Kruschke (2014) Chapter 11.
[8] Kruschke (2014) p 25.
[9] Wikipedia: Stephen Fienberg (9.5.2022).
[10] Wikipedia: Thomas Bayes (7.5.2022).
[11] Wikipedia: Pierre-Simon Laplace (7.5.2022).

inventions of likelihood, hypothesis testing and confidence intervals. The invention of those methods is said to revolutionized the use of statistics in the field of science. And those methods are still used in modern statistics.[12]

Nevertheless, the idea of inverse probability – or subjective probability – was not forgotten and it was developed in parallel with the "mainstream" frequentistic approaches. During the second world war statistics were seen important component in the science supporting war effort, so statistical research was funded significantly. Statistics was developed fundamentally during war and post war periods, however at that time the frequentistic approach was undoubtedly on the forefront.

Considering Bayesian approaches, one of the most important developments were the ideas behind *Monte Carlo* methods and the development of *Metropolis* algorithm. Both of them introduced in the Manhattan project, which worked towards the development of an atomic bomb.[13]

Instead of analytical mathematical methods, Monte Carlo methods use randomly generated samples (simulated data) to track statistical behavior of a population (or phenomenon). The name Monte Carlo comes is an analog to the random events happening inside a casino.[14] One of the most simplest examples of Monte Carlo is counting the approximation of Pi by drawing a square and a circle inside the square. Then adding dots to random coordinates on that square. Lastly counting which of those dots are inside the circle, multiplying the number of these dots by four and dividing this by the number of all the dots in the square.[15] Metropolis algorithms are an algorithms used to generate (approximate) the statistical behavior of population or phenomenon (the posterior distribution).[16]

In other words: Monte Carlo methods are used when the exact mathematical values (analytical mathematics) are not known and have to be approximated with random numbers. Metropolis algorithms are a type of algorithms which generates random values and follows those mathematical patterns (functions or distributions). So the behavior of the function or distribution is kind of "drawn" with the Metropolis algorithm.

In the latter half of 20th century there was a newly spawned interest towards Bayesian statistic. Fienberg calls it "Neo-Bayesian revival". One of the notable developers of the

[12]Fienberg (2006) p 1 - 9.
[13]Fienberg (2006) p 9 - 14.
[14]Mooney (1997) p 1 - 3, Kruschke (2014) p 144 - 145.
[15]This method is almost over simplified, but may give some kind of clue about the basics of Monte Carlo. It was a demo exercise for the algorithm course in the department of information technology.
[16]Kruschke (2014) p 144 - 145.

approach was a classical historian Robert Schlaife.[17] According to Fienbergs interview with Schlaifes colleague Howard Raiffa:

> when he [Schlaife] was asked to teach statistics, 'he read [frequentistic statistics] Fisher, Neyman and Pearson ... and he concluded that standard statistical pedagogy did not address the main problem of a businessman: how to make decisions under uncertainty. Not knowing anything about the subjective/objective philosophical divide he threw away the [frequentistic] books and invented Bayesian decision theory from scratch'[18]

In the late 1980s an important event was the popularization Monte Carlo *Markov chain* methods (MCMC).[19] Markov chains are mathematical systems which represent processes as chains. They basically have a state which changes according to known rules.[20] Also the overall increase in the computation capacity has made Bayesian methods more accessible. According to Kruschke:

> It is MCMC algorithms and software, along with fast computer hardware, that allow us to do Bayesian data analysis for realistic applications that would have been effectively impossible 30 years ago.[21]

All in all, the idea of inverse probability and Bayesian statistics is nothing new. And to a historian working daily with uncertainty, the main idea seems quite intuitive. And as I will discuss below, the approach has already been applied to historical data. However, what comes to the historiography of Bayesian statistics, Fienberg mentioned that to cover the subject comprehensively would take a whole book.[22] Maybe this could be a task for some science historian.

2 Bayesian approaches to history

What comes to the problems with Bayesian statistics. The first problem is shared by all of the statistical – or even qualitative – methods. If the sources and data is bad or corrupted no method – no matter how elaborated – can fix it. The problem is so-called

[17] Fienberg (2006) p 15 - 20.
[18] Fienberg (2006) p 17 - 18.
[19] Fienberg (2006) p 27
[20] Brilliant: Markov chains (9.5.2022).
[21] Kruschke (2014) p 144.
[22] Fienberg (2006) p 3.

"garbage in garbage out" (GIGO).[23] This was also mentioned by historian Jo Guldi in her lecture about the computational approaches to historical texts.[24]

The next problem is that Bayesian models are complex and they require a certain amount of mathematical understanding and programming skills. Of course, achieving those skills, will take time and effort. The most common software for these tasks are *R* with *Stan*. R is a programming language designed specially for statistical computation and graphics. It was first released in 1995.[25] Stan is a software package which basically runs the Monte Carlo-methods in the modeling process.[26]

But is the imperative need to learn the mandatory mathematical and programming skills inherently a bad thing? In his book about statistical inference and scientific argumentation Mikko Ketokivi is concerned about the side effects that easy-to-use statistical calculation programs pose. With these programs scholars may use complex statistical models they do not understand in the first place. He also mentioned the enormous business behind these programs.[27] For me using methods one does not understand, and drawing conclusions based on that, sounds more of a iffy research than the democratization of statistics.

As I already mentioned, the models are complex so they need a longer computing time. As Guldi mentioned in her lecture about text mining that, TF-IDF (term frequency - inverse document frequency) algorithms takes approximately 10 minutes to run in a computer.[28] More complex Bayesian models can take hours or even days to calculate. A case in point, the historical census calculated by Voutilainen et al took computational time of 144 hours.[29]

Overall, the more computational time takes more resources and funding. For more resource intense Bayesian models an external computer may be needed to be rented. And human mistakes like typos or calculation errors or small careless errors may get expensive.

Lastly there is a fundamental question of what we can infer with quantitative methods. As Kustaa H. J. Vilkuna mentioned in his lecture about the 20th century English historiography that, for example a left wing historian E. P. Thompson (3.2.1924 – 28.8.1993)[30]

[23]Cambridge dictionary: garbage in garbage out (8.5.2022).
[24]Guldi (3.5.2022).
[25]HOPL: R (9.5.2022).
[26]Kruschke (2014) p 399 - 400.
[27]Ketokivi (2015) 7.7 Haaste 7: Käyttäjäystävälliset tilasto-ohjelmat.
[28]Guldi (5.5.2022), tfidf (8.5.2022).
[29]Voutilainen et al (2020) p 1183.
[30]Wikipedia: E. P. Thompson (8.5.2022).

criticized statistical approaches to history. According to Thompson numbers do not describe how historical people experienced their living standards, work life or poverty.[31]

This critique is relevant and quantitative methods can be applied to different questions than qualitative methods. One can not interpret a historical person's feelings or experiences through numbers. However, this goes to the other way too. One can not for example estimate regional mortal rates in the Finnish 1860s famine from some individuals private exchange of letters.[32] Next I am going to discuss the pros and possibilities of Bayesian statistic in the historical research.

In fact, Bayesian approaches to historical sources is nothing new. One of the first high profile Bayesian project was a historical one. In the summer of 1959 a group of statisticians used Bayesian approaches to determinate who had written the 1780s Federalist papers in America. The work was finished in 1963 and it is considered the first empirical use of Bayesian methods with considerable use of computers.[33] Also Guldi mentioned this project in her lecture, but she emphasized more the usage of text data than the Bayesian core of the actual statistical model.[34]

The article that inspired this essay is instead a historical census with Bayesian approach. In their article *A Bayesian Reconstruction of a Historical Population in Finland, 1647–1850* Miikka Voutilainen, Jouni Helske and Harri Högmander are re-evaluating Finnish population between 1647 and 1850 using Bayesian approach. They have used historical records such as baptists and perished records as their data. With their approach they got some different results than previous population historians such as Eino Jutikkala.[35]

Another possibility for Bayes statistics could be a following kind of question. A few years ago I was talking in our campus with a scholar part taking in the *Contextualizing Finnish Early Modern Economy (1500-1860): Construction and Analysis of Aggregate Time Series* -project (01.09.2017 - 31.08.2021)[36]. She was concerned how she could get women's work represented to the statistics, because there is no records of women's work. I thought that Bayesian approach could work on questions like these.

I assume, we *a priori* know that women have worked (at least in the households). Secondly we know that the records about men's work are not comprehensive either. So

[31] Vilkuna (20.4.2022).
[32] See for example Miikka Voutilainen (2016): Poverty, inequality and the Finnish 1860s famine.
[33] Fienberg (2006) p 23 - 24.
[34] Guldi (5.5.2022).
[35] Voutilainen et al (2020).
[36] juy:Contextualizing Finnish Early Modern Economy (1500-1860): Construction and Analysis of Aggregate Time Series (9.5.2022).

with careful examination of the existing sources and more historical prior knowledge it would be possible to "fill the gaps" create prior distributions about women's and men's work. Then use computational methods to approximate the data.

Besides these examples, Bayesian statistics has already been used in historical research. Overall, the Bayesian approach is not in the "marginal" of statistics and it is commonly used in mundane applications we use for example weather forecasts or machine learning. Also the newest climate model *Globally resolved surface temperatures since the Last Glacial Maximum* written by Osman et al in 2021 is based on Bayesian modeling.[37]

What comes to my own research. At the moment, I have written a few snippets with R and only downloaded Stan to my computer. And as mentioned above, I have only a basic conceptual understanding of Bayesian approach. So, I would be in no means ready to write a complete Bayesian model. However, the approach and its possibilities interests me and after some practice it would be possible to apply.

Sources

Brilliant: Markov chains: `https://brilliant.org/wiki/markov-chains/` (checked 9.5.2022).

Cambridge dictionary: garbage in garbage out: `https://dictionary.cambridge.org/dictionary/english/garbage-in-garbage-out?q=garbage+in%2C+garbage+out` (checked 8.5.2022).

Fienberg, Stephen E. "When did Bayesian inference become" Bayesian"?." Bayesian analysis 1.1 (2006): 1-40.

Gelman, Andrew, John B. Carlin, Hal S. Stern, David B. Dunson, Aki Vehtari, ja Donald B. Rubin. Bayesian Data Analysis. Third edition. Boca Raton: CRC Press, 2014.

Guldi, Jo: lecture about computational approaches to historical texts 3.5.2022, University of Jyväskylä.

Guldi, Jo: lecture about text mining 5.5.2022, University of Jyväskylä.

HOPL: R: `https://hopl.info/showlanguage.prx?exp=2426&language=R` (checked 9.5.2022).

[37]nature: Globally resolved surface temperatures since the Last Glacial Maximum (10.5.2022).

Jyu: Contextualizing Finnish Early Modern Economy (1500-1860): Construction and Analysis of Aggregate Time Series: https://www.jyu.fi/hytk/fi/laitokset/hela/tutkimus/keskittymat/growth (checked 9.5.2022).

Ketokivi, Mikko. Tilastollinen Päättely Ja Tieteellinen Argumentointi. Helsinki: Gaudeamus Helsinki University Press, 2015.

Kruschke, John. Doing Bayesian Data Analysis, 2nd Edition. Academic Press, 2014.

Matematiikkapakki, todennäköisyyslaskenta: https://tim.jyu.fi/view/kurssit/matematiikka/matikkapakki/todennakoisyyslaskenta2/kombinatoriikka-ja-todennakoisyys#klassinen-todenn%C3%A4k%C3%B6isyys (checked 9.5.2022).

Mooney, Christopher Z. Monte Carlo Simulation. Thousand Oaks, Calif. ; London: SAGE, 1997.

Nature: Globally resolved surface temperatures since the Last Glacial Maximum: https://www.nature.com/articles/s41586-021-03984-4 (checked 10.5.2022).

tfidf: http://www.tfidf.com/ (checked 8.5.2022).

Vilkuna, Kustaa H. J.: lecture about 20th century historiography 20.4.2022, University of Jyväskylä.

Voutilainen, Miikka; Helske, Jouni; Högmander, Harri: "A Bayesian Reconstruction of a Historical Population in Finland, 1647–1850". Demography 1 June 2020; 57 (3): 1171–1192. doi:https://doi.org/10.1007/s13524-020-00889-1

Wikipedia: E. P. Thompson: https://en.wikipedia.org/wiki/E._P._Thompson (checked 8.5.2022).

Wikipedia: Pierre-Simon Laplace: https://en.wikipedia.org/wiki/Pierre-Simon_Laplace (checked 7.5.2022).

Wikipedia: Stephen Fienberg: https://en.wikipedia.org/wiki/Stephen_Fienberg (checked 9.5.2022).

Wikipedia: Thomas Bayes: `https://en.wikipedia.org/wiki/Thomas_Bayes` (checked 7.5.2022).

Further reading

- Kruschke, John. Doing Bayesian Data Analysis, 2nd Edition. Academic Press, 2014.
- Töttö, Pertti. Paljonko on Paljon?: Luvuilla Argumentoinnista Empiirisessä Tutkimuksessa. Tampere: Vastapaino, 2012.
- Ketokivi, Mikko. Tilastollinen Päättely Ja Tieteellinen Argumentointi. Helsinki: Gaudeamus Helsinki University Press, 2015.