

Bayes and i: a gentle introduction to the bayesian approach

Abstract

This article describes the concept of the Bayesian approach in a very easy way. We expect that this simple and straight commentary would help a lot of those who have felt difficulty in embracing the Bayesian paradigm.

Volume 5 Issue 2 - 2017

Heon-Jae Jeong,¹ Wui-Chiang Lee²

¹The Care Quality Research Group, Chuncheon, Korea

²Department of Medical Affairs and Planning, Taipei Veterans General Hospital & National Yang-Ming University School of Medicine, Taipei, Taiwan

Correspondence: Wui-Chiang Lee, Department of Medical Affairs and Planning, Taipei Veterans General Hospital & National Yang-Ming University School of Medicine, Taipei, Taiwan, Tel +886-2-28757120, Fax +886-2-28757200, Email leewuichiang@gmail.com

Received: February 13, 2017 | **Published:** February 17, 2017

Disclaimer

Before getting started, we want to make clear that this short article is a highly simplified introduction for researchers unfamiliar with the Bayesian approach. No mathematical formulae will be presented. Our intention is just to convey the basic concepts in an easily digestible format, using a slight modification of Hoff's work as an example (2009).¹

What we habitually do when analyzing data: the frequentist approach

Assume that we are seeking a parameter of a population, such as the mean weight of teenagers in a city. We begin by selecting a representative sample that consists of a certain number of teenagers, and calculate the mean of their weights; then considering that we did not measure all the teenagers in town, we calculate a 95% confidence interval of the mean we obtained. It is simple and clear.

Yet, behind the scenes, a very strong assumption—or belief—resides: that there is a fixed single population parameter. In this case, the mean value of weights for teenagers in the city is a single value. Therefore, without a serious flaw in the sampling process, the parameter is **constant**, though unknown, across all possible samples.² However, what if the parameter is not fixed but rather has a distribution? This is where the Bayesian approach comes in.

Re-wiring our brain: the bayesian approach

Before diving into the Bayesian approach, it is worth thinking about how our knowledge generally evolves. Choose whatever topic you like, and search articles about it from the past ten years. Focus on their references, and you will certainly find that newer researches are almost always based on—or at least influenced by—the previous studies, leading to new results. Then, even newer research builds upon the basis of this evolved knowledge; indeed, this is how knowledge continuously advances. The Bayesian approach can be understood in the same way: there is prior, existing information, and new data are applied to it, resulting in updated information. Now, let us bring this

simple idea into the statistical arena by modifying a very practical example described in Hoff (2009, 3).¹

Assume that there is an issue about a disease in a country, so your city decides to measure its prevalence, and asks you to conduct a study. Due to budget constraints, you choose a small number of citizens (sampling) and conduct a test. Surprisingly, the results show that no one has the disease. From the frequentist perspective, the prevalence in the sample is zero. Thus, both the upper and lower bounds of the 95% confidence interval are also zero; however, the result is not persuasive at all, since the disease is already a national issue—even your neighbor has it. People, including the study's sponsor, start raising questions about the precision of your study. At this point, you may feel tempted to say that the sample was peculiar although the sampling process was correct. However, once you say so, you lose ground to justify the results from any potential sample. For example, how do we know if a certain sample is appropriate (not peculiar); a sample with one patient or two patients? There is no answer.

Now, we apply the Bayesian approach to the same sample with zero patients found. While reading the following description, please refer to Figure 1 (i.e., the overall shape of the lines). This time, we decide to take advantage of the already existing information that “the **prevalence** of the disease in similar cities of the country **ranges** from 0.05 to 0.20, an average of which is **0.10**” (dotted line in Figure 1). Here is the most important concept of this article: the *parameter, which is the prevalence in this example, is not fixed, but has a distribution*. We call this information *prior*—technically speaking, *prior distribution*. Then, to the prior distribution of the parameter, we apply the information from the sample, the zero prevalence, and get an updated distribution of the prevalence with a mean of around 0.05 (solid line in Figure 1). We call this updated distribution of the parameter the *posterior distribution*. How to conduct these, including the choice of distribution is beyond the scope of this article; however, the take-home message is that what is shown in Figure 1 is a *distribution of parameter*, a concept that contradicts the frequentist's axiom: the parameter is fixed and constant.

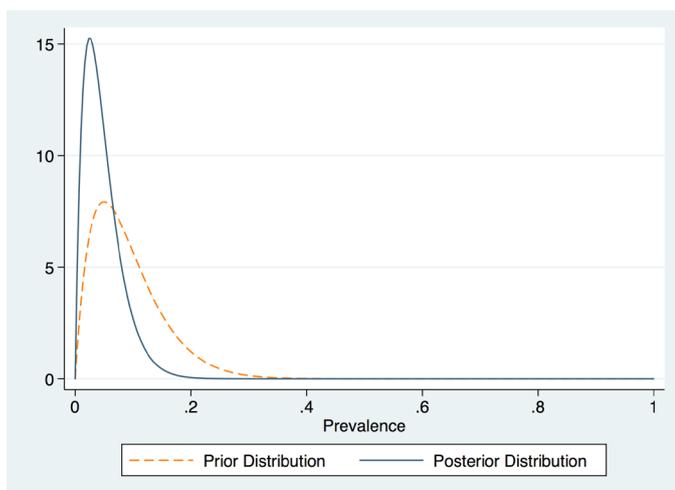


Figure 1 Prior and Posterior Distributions of Prevalence (parameter of interest in this example).³

You might argue that this approach is too much of a stretch, asking “on what grounds can we change the result from the sample at hand?” We understand your reservations. Let us summarize the above example in a more practical way: “The posterior information is a compromise between the prior information and the sample at hand.” The direction and amount of the compromise depend on the precision of the sample, which is largely determined by sample size. To illustrate, if the sample size in the above example was only around 10, the results from the sample may not be very precise (unreliable); thus, the impact of the sample result will be downplayed, causing the posterior distribution to look very similar to the prior distribution. Simply put, only a negligible update is made by the sample data. On the other hand, if the sample size is one thousand, then we can expect much higher precision of the results from the sample. Then, the impact of prior distribution will be minimized, sometimes to negligible levels. Thus, the posterior distribution will be greatly shifted from its original prior to the current result from the sample at hand, meaning that a major update of information occurs. In this situation, the posterior information will be almost the same as from the frequentist perspective. In this respect, some researchers jokingly call the Bayesian (especially, the empirical Bayesian) approach a buffer or safeguard using prior information: if current data are weak, then it gets more help from prior; if current data are strong, then the current explains almost everything with minimal help from the prior. Simply put, Bayesian is flexible and robust.

If you are hooked, then let us move on to the real beauty of the Bayesian approach, which can be found in the credible interval, the counterpart of the confidence interval from the frequentist inference. The typical 95% confidence interval from the frequentist perspective means that if the same sampling process takes place over and over and the confidence interval is calculated for the corresponding sample, then 95% of those intervals may contain the fixed, constant, or so-called “true” parameter of interest. Thus, strictly speaking, what we can say from this is only a *plausible range of the parameter*. By contrast, a 95% credible interval from Bayesian inference can provide us with a *range for a parameter where the probability that the parameter lies within that range is 95%*.³ Interestingly, many frequentist perspective-based researchers interpret the confidence interval in the latter way, but that is a huge mistake. Such probability can only be obtained through Bayesian inference. (You may want to combine Figure 1 and

the concept of Area under Curve.) If we expand this idea beyond the credible interval, we can even provide information such as “there is X% probability that the prevalence of disease in the city is less than 0.10.” This form of information is direly needed in policy decision making and resource allocation, but can never be reported by the frequentist approach, which provides only a single parameter estimate and its confidence interval—a *plausible range*.

Do we really need to understand and use bayesian analysis?

Many of you with only frequentist analyses techniques at hand may be confused and feel the desire to disavow yourselves of the Bayesian approach. This is completely understandable; indeed, admitting that a parameter has a distribution is a real paradigm shift, and requires a huge leap of faith, probably as much as was needed when science leapt from Newtonian dynamics to Einstein’s Theory of Relativity.

Despite the difficulty of making this shift, we strongly recommend that readers consider embracing the Bayesian approach. One important reason is that many analysis methodologies in your current arsenal already utilize the Bayesian approach behind the scenes. Assume that we are conducting a meta-analysis of 50 published articles. Though there are complex statistical notations, the essential idea of meta-analysis is to consider the result of each article as a parameter, and find a distribution of those 50 parameters. This is basically the same as what we saw in the above example of disease prevalence, but now each city is an article and prevalence of disease in each city is the result of each study.

Another very common use of the Bayesian framework is multilevel analysis. When data are clustered, the canonical assumption of the typical regression model—namely, that “observations (technically, residuals) are independent among each other”—does not hold. Then we apply the Bayesian way of thinking, treating each cluster like a city in the above example. Although we are running multilevel models on our computer just like typical frequentist regression models, the Bayesian concept is actually functioning behind the scenes. In addition, we can naturally understand that a longitudinal analysis may rely on the Bayesian paradigm. As we mentioned previously, the Bayesian approach basically means updating prior information to get newer information.

The bottom line is that we have already been using several methodologies in which Bayesian analysis is already built in. Thus, the question, “Do we have to use Bayesian?” might be just an oxymoron.

Final comments, promises and requests

There are many more examples, but we will stop here. We believe this short article already provides enough insight into Bayesian for those unaccustomed to it. The authors will demonstrate the full application of the Bayesian approach in articles dealing with the rates of hospital-wide or even national-wide patient safety-related adverse events, healthcare professionals’ patient safety culture levels across a country and its changes, and many other clinical and cultural issues. In those articles, we promise to provide detailed mathematical descriptions, including simulations, which are an essential component of fully utilizing Bayesian analysis. Please open your mind to a Bayesian approach. If possible, please come up with questions. Perhaps, you’re wondering what can be done if there is no clear prior information? If you feel such curiosity, yes, you’re already in. We will provide an answer.

Acknowledgement

None.

Conflict of interest

None.

References

1. Peter HD. A first course in Bayesian statistical methods. *Springer Science & Business Media*. 2009.
2. Scott ML. Introduction to applied Bayesian statistics and estimation for social scientists. *Springer Science & Business Media*. 2007.
3. StataCorp. Stata Bayesian Analysis Reference Manual: Release 14. College Station, TX: Stata Press. 2015.