2023 ■ Vol. 19 ■ no. 3

The Bayesian Approach is Intuitive Conditionally to Prior Exposition to These Examples

Philippe Pétrin-Pomerleau [a], Coralie Vincent [b], Paola Michelle Garcia Mairena [a], Ece Yilmaz [a], Annie Théberge Charbonneau [a], Tracy Husereau [a], Ghislaine Niyonkuru [a] & Grace Jacob [a]

[a]University of Ottawa, School of Psychology
[b]University of Ottawa, Interdisciplinary School of Health Sciences

Abstract ■ There is a range of statistical approaches available to researchers. Nevertheless, in the probabilistic context, the frequentist approach is dominant, from the scientific literature to the teaching of statistical methods in higher education institutions. However, research questions are diverse, and other probabilistic statistical approaches may be advantageous in specific contexts. The methods used by researchers are derived mainly from their training. Unfortunately, alternative approaches, such as the Bayesian approach, are rarely taught, which may, in part, be due to the complexity of teaching them. This article aims to address this problem by presenting a series of fictitious examples illustrating the concepts behind Bayesian reasoning. It is intended as a tool for novice researchers looking to gain a basic understanding of the Bayesian approach. The prior, likelihood and posterior concepts will be illustrated by scenarios that learners can identify with. It is expected that novice researchers who have internalized the concepts of the Bayesian method, partly through these intuitive examples, would be more inclined to learn about this alternative statistical approach and consider using it in their research field. This could, in turn, help diversify the statistical methods used throughout the scientific literature.

Keywords ■ Bayesian; Priors; Statistics Education; Vulgarization; Meta-statistics.

Acting Editor ■ Denis Cousineau (Université d'Ottawa)

ppetr045@uottawa.ca

10.20982/tqmp.19.3.p244

Introduction

Statistics are used to make sense of uncertain information (Cressie & Wikle, 2011, p. 4). If the results were undeniable, statistics would not be necessary. Therefore, the results have a certain level of uncertainty, and you can only have a certain level of confidence in them (Lele, 2020; Tukey, 1991; Zyphur & Oswald, 2015). This opens the possibility for different approaches to address the uncertainty differently, which can lead to different conclusions about the data. We frequently make decisions based on varying amounts of information in various contexts, and good decision-making requires using the most appropriate approach for the situation. There is extensive literature on the differences between the two main probabilistic approaches: the frequentist and Bayesian approaches (Bandyopadhyay, 2011; van de Schoot et al., 2017). However, the frequentist approach dominates scientific literature and education curricula (Bland & Altman, 1998; Dogucu & Hu, 2022; van de Schoot et al., 2017). This prevalence of one approach over the other could be explained by the limited resources available to teach and promote the understanding of the Bayesian approach (Lecoutre, 2006). The purpose of this study is not to debate which approach is best but rather to provide an additional resource for students looking to be introduced to Bayesian reasoning or for educators looking for a resource to assist them in teaching it. Various examples related to different domains of everyday life will be presented and explained to facilitate the understanding of the basic concepts behind the Bayesian approach. Some concepts are simplified to facilitate the integration and contrasts are dichotomized to ease the reader into the possible issues of the Bayesian approach. The present article will focus on a narrower and more intuitive difference between the Bayesian and the frequentist approaches, namely, the inclusion of prior knowledge into the statistical analyses.

Frequentist approach

The frequentist approach does not include preliminary information in the analyses, as it exclusively uses the data gathered during the experiment to arrive at conclusions (Jaynes, 1976; van de Schoot & Depaoli, 2014). This often involves comparing the data to a null hypothesis and looking at a p-value to conclude on the level of certainty that the effect in the data exists (Dancey & Reidy, 2007). In this paradigm, sufficiently large and representative data samples are necessary to reduce the uncertainty to an acceptable threshold (Faber & Fonseca, 2014). This approach is widely used in scientific research as it allows researchers to carefully design their experiment to meet the requirements for the data (e.g., a sufficient number of participants and unbiased collection methods).

However, this approach can fail when the conditions are unmet, as any problem with the data could lead to a problem in forming the conclusion (Cohen, 1994; Jaynes, 1976; Lecoutre et al., 2003). For example, imagine that you want to know the probability of rolling any number on a die and try to figure it out by rolling it only five times. The frequentist's conclusion, entirely derived from the data, could lead you to believe that it is impossible to roll a three simply because it did not happen during your experimentation. This would be an erroneous conclusion because of the insufficient size of the data. If you roll it 100,000 times and never roll a 3, you could conclude that the die is probably rigged.

Although no prior information is used in the statistical analyses, past knowledge is still present under the frequentist methodologies, but they are not mathematically quantified (Kaplan, 2014). For example, variables are selected to be measured before any scientific experiment. These decisions are based on the researcher's prior knowledge and assumptions that, for example, anxiety will be associated with depression. The variables studied are not chosen at random, and there is a high probability of them being relevant before doing the study. Other examples are assumptions about the structure of the statistical model, such as in mediation (MacKinnon et al., 2007). For instance, a researcher could use information from a literature review to hypothesize that attention would mediate the association between sleep quality and academic achievement.

Similarly, confirmatory factor analysis sets the structure and number of factors before collecting the data and running the studies (Bollen, 1989). These assumptions rest on prior information possessed by the researcher. Still, they are not quantified in the analyses, even if they are more likely to be of interest than other random variables.

Bayesian approach

The Bayesian approach adds to the data by considering this prior knowledge in the statistical framework to arrive at a conclusion that combines the data and the preceding information (Kurt, 2019; van de Schoot & Depaoli, 2014). Going back to the die example: if you enter as prior knowledge that each side has a 1/6 chance of being rolled and then experiment with five random rolls without ever getting a 3, the conclusion will not be that it is impossible to roll a 3, even if you don't roll any 3's, but rather that the probability is less than 1/6. We arrive at this conclusion by starting with the prior knowledge (1/6 chance) and adding the data (zero chances) to conclude that it is less likely to be 1/6 after the experiment, but not much less because of the low number of rolls. This example illustrates how, with a lack of data or a biased sample, the Bayesian approach will provide a less extreme value because it is balanced with prior knowledge (Johnson et al., 2022). If we roll the die an infinite number of times, the two methods will arrive at the same conclusion: that a single die has a 1/6 chance of rolling any number.

This fictitious situation highlights a condition for which the Bayesian approach would be vastly superior to the frequentist approach, and situations as extreme as this are rare in real life. To address scientific questions, the Bayesian approach is not always ideal, and when it is, the difference is not as extreme as in our example (Hackenberger, 2019). Nonetheless, the Bayesian approach to probability can be helpful in many situations (Winkler, 2001), and we would benefit from it being introduced to future researchers (Depaoli et al., 2017; Beard & West, 2017).

Although this approach may appear more complex initially, its reasoning applies to numerous real-life situations (Johnson et al., 2022). Humans are often expected to make decisions or form opinions based on uncertain conditions and small amounts of data or experiences (Kurt, 2019). As such, this article will bring forward the main concepts surrounding the inclusion of previous knowledge in the Bayesian approach using real-life examples that illustrate Bayesian reasoning in an intuitive and accessible way. We believe this approach can be intuitive to novice researchers and students conditionally on it being properly presented.

Prior, likelihood, and posterior

We have been referring to these three components of the Bayesian framework without directly naming them. Still, we should define them before diving further into how they interact. The prior refers to the knowledge or opinion acquired before the data is collected (Johnson et al., 2022). Theoretically, a prior could be constructed from the past collected information, scientific knowledge, or hearsay (Kurt, 2019). For example, if we want to know if we will enjoy reading a specific book, we could judge it by its cover (information available before reading). This can be considered an example of a weak or uninformative prior, as the cover does not necessarily provide much information on the book's contents (Depaoli et al., 2017). A more substantial source of preliminary information could be to have read other books by the same author or ask friends' opinions who have already read this book. Then, reading parts of the book could be considered as the data collection, which, in statistical terms, is referred to as the likelihood (Depaoli et al., 2017; Kurt, 2019). In this example, reading the entire book would mean having all the necessary information to come to a conclusion, suggesting that statistics would not add anything useful to form your opinion on the book.

However, let's say that you read the book's first half and want to decide if you should finish it. This decision would be based on a combination of all the priors, which could have included the cover's look, your opinion about other books by the same author, and your friends' opinions about the entire book. It would also be based on the likelihood, which would be the data you collected or the idea you formed while reading the book's first half. The probability of you finishing the book is called the posterior in the Bayesian framework and would be based on a combination of the priors and likelihood (Kurt, 2019). These three components can have several characteristics and interact in different ways to give rise to various situations described in the following sections.

Single versus multiple priors

To begin, priors are the previous knowledge of the situation (Depaoli et al., 2017). They can come from multiple sources, have different conclusions, or relate to other parts of the circumstances (Epstein & Schneider, 2007). They can have different weights and contributions to that opinion and are typically categorized as uninformative, empirical or informative priors (Zyphur & Oswald, 2015).

Here is an example we will expand on over the following few sections: you are going on a blind date with someone your friend wants you to meet. Before you meet that person and find out how they are, you may receive hints about their personality from different sources. You know that their astrology sign is Libra (Prior 1), that you both like the same type of music (Prior 2) and that your friend who knows that person thinks you will get along (Prior 3). These different sources of information contribute differently to your prediction of how well you will get along with this person. You could have only one relevant prior or add priors you consider appropriate, but more priors are not necessarily better and could bias your conclusion, which we will address later. To best assess if you will get along with that person, the likelihood (meeting them) will contribute significantly, but one date might not be sufficient. Your conclusion may need to be updated over time, meaning your first date could become an additional prior for your next meeting.

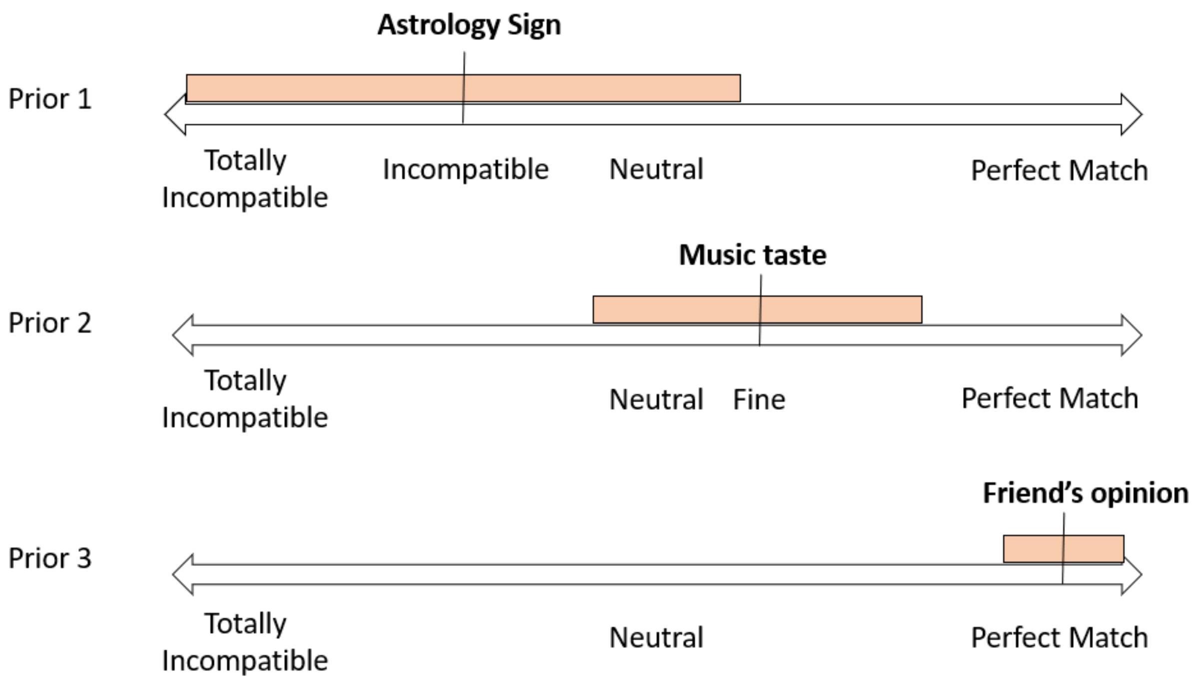
Expected value and variance

Priors contribute differently to forming your opinion because of how strong or weak of a predictor each source of information is for how well you will get along with your date (Depaoli et al., 2017). Weak priors are labelled uninformative priors in the Bayesian literature (Zyphur & Oswald, 2015) because they do not bring much information or information with low certainty, in opposition to strong prior (informative priors). The priors' quality can be described according to the expected value and the variance, or in other words, the claim made before data collection and how confident you are with that prediction (Depaoli et al., 2017). The prior has expectations of what will be observed and assigns a value to it, but it also comes with variance around this value. For example, in the dating scenario, the fact that you like the same type of music as your future date could predict that you should "get along fine," whereas your friend's opinion is that you two would be a "perfect match." The strength of your friend's claim is more potent by using the words "perfect match" than the claim of "getting along fine" from having similar musical tastes; therefore, the expected value is higher. Their astrology sign may say you are "incompatible," which is the opposite claim from the music preference and your friend's opinion. This astrology claim could be stronger than the musical preference claim but weaker than your friend's statement (see Figure 1). The variance, or your confidence in this claim, is the other characteristic that will affect your decision (Zyphur & Oswald, 2015). It refers to the degree of precision or confidence in your claim or, in other words, how many alternatives could realistically exist (Depaoli et al., 2017). In our example, the astrology sign has a strong assertion in one direction but such a high variance that it would be possible to have a result on the positive side.

The claim from your friend's statement is much more credible, as that person knows both you and your date very well, hence the smaller variance. You could choose the prior you believe to be the best predictor, or they can all be combined to form your prior, which in this case, would be your overall opinion of how you will get along with this person before meeting them. This example is represented in Figure 1.

The expected value and variance also apply to the likelihood or in other words the data (Zyphur & Oswald, 2015). Your data collected could identify a strong or a weak claim,

Figure 1 ■ Visual representation of multiple priors, their expected value and variance. The double-sided arrow represents the scale of possible claims on how compatible you are with your blind date, ranging from "totally incompatible" to "perfect match." The black vertical line represents the value of the claim or expected value, and the further towards the extremities, the stronger the claim is. The orange rectangle represents the variance of the claim. The rectangle covers the values that could happen; therefore, the smaller the area, the smaller the claim's variance.



with low or high variance. For example, when meeting the person, you could feel like it went well (strong claim) or just fine (weak claim), and this could be based on a five-minute conversation (high variance), a five-hour date (medium variance), or years of interactions (low variance). It is the same for the conclusion (posterior), such that it can have a strong claim if you have a substantial prior and strong likelihood or a weak conclusion if you have weak prior and flimsy data. Nonetheless, as seen in the examples below, many possible combinations of strength levels for the prior and likelihood can affect the posterior in varying ways.

Aligned or opposed prior and likelihood

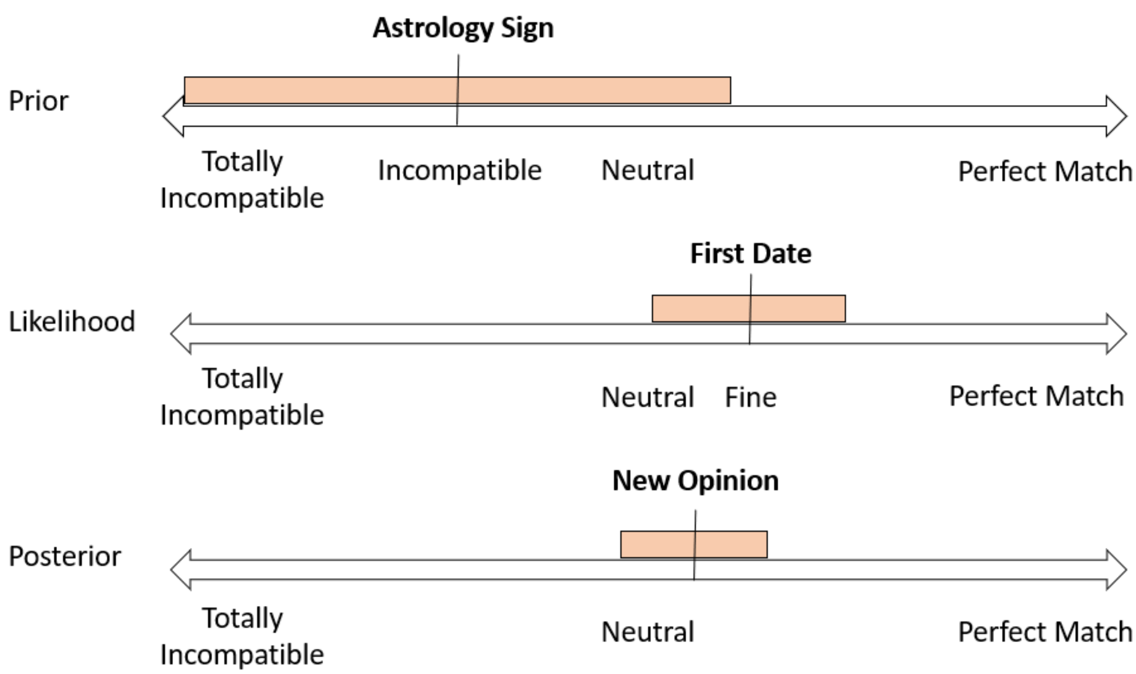
Another essential characteristic of the prior and likelihood is the direction of their claims. The prior and the likelihood can be aligned, which means that the direction of their respectful claims is the same (Reimherr et al., 2014). For example, your friend says you will get along with your date (positive prior), and then you do get along (positive likelihood). Contrastingly, they could be opposed (prior-likelihood dissonance), such that your music taste predicted that you would get along fine (positive prior), but the date goes horribly (negative likelihood; Reimherr et al., 2014). In any case, the posterior will be a compromise between the claims of the prior and the likelihood. It will be aligned or opposed to the priors and the data depending on the characteristics of these components (Depaoli et al., 2017). This will be explored further in the following sections.

Changed, confirmed, or softened belief

As stated before, the posterior combines the information provided by the prior and the likelihood (Depaoli et al., 2017). It combines the prediction with the collected data while considering their expected value, variance, and direction. As visualized in Figure 2, the result will be between the two values but dragged more toward the claim with the lowest variance (Zyphur & Oswald, 2015).

For example, the prior from your astrology sign, which has a high variance and is considered less informative, will have less impact than meeting the person. As a result, the posterior will be closer to the data than the prior. It could

Figure 2 ■ Visual representation of the combination of the prior and likelihood into a posterior. The double-sided arrow represents the scale of possible claims on your compatibility with the blind date ranging from "totally incompatible" to "perfect match." The black vertical line represents the strength of the claim, and the further towards the extremities, the stronger the claim is. The orange rectangle represents the variance. The rectangle covers the values that could happen; therefore, the smaller the area, the more confidence you have about this claim. The third arrow represents the posterior; the value is between the prior and the likelihood values, but more towards the likelihood because of its lower variance.



also be the opposite: if your friend is sure that you will be a perfect match, but after a five-minute conversation, you don't agree, they may tell you to "Give it more time to get to know each other." In this case, maybe you would give it another try because your conclusion may be closer to your friend's claim than the short experience you had with your blind date.

The relationship between the prior (previous belief) and the posterior (new belief) depends on the data (Zyphur & Oswald, 2015). Your future beliefs can change if the prior and the likelihood are in different directions and the data is stronger or has less variance (Depaoli et al., 2017). For example, maybe you thought you would not like your future date because they are Libra, but after meeting them, you changed your mind and really liked them. On the other hand, your belief can be confirmed and strengthened if both the prior and the likelihood are in the same direction; they give you a posterior with less variance than each claim separately (van de Schoot & Depaoli, 2014). This expresses the phenomenon of accumulating evidence, where the more experiences provide you with the same conclusion, the more confident you are. This reinforces or confirms your belief, which can become your prior for a future encounter with data and make it harder to change your views as you gain confidence in your result over time and see less possible variance around that value (Kurt, 2019). New data will need to be stronger in the opposite direction to change your beliefs (Depaoli et al., 2017). Finally, a belief can be softened (Kurt, 2019). If you have a prior stronger than the data and are in opposite directions, your belief will remain closer to the prior, although not identical. It may become a bit closer to the data, and you might slowly get convinced by evidence, but you would require additional data to change direction.

Understanding Bayesian Reasoning Using Real-life Scenarios

In this next section, real-life scenarios exemplify Bayesian reasoning and the concepts of prior and likelihood. They illustrate how their respective strength, variance and direction may influence how they interact to form the posterior. Examples will be presented by increasing the order of complexity, beginning with a conceptual example and slowly introducing numbers and basic probabilities.

Example 1

> *Since moving to the countryside, a long time ago, your aunt has told a story about seeing a spaceship right over her cornfield. She saw it in broad daylight and called her husband to show him. Unfortunately, the ship was gone when they returned to the field.*

In this scenario, you are trying to determine whether or not you believe in aliens. The first information you are using to base this opinion on is the story your aunt has told you, in which she says she has seen a spaceship in broad daylight. This would be considered a weak prior because it comes from hearsay, which is not a reliable source of information. Let's add components to our example:

> *You recently came across a public forum on the Internet with a hundred similar testimonies to your aunt's. People tell of similar events in different circumstances and parts of the world.*

The added information about the peoples' testimonies in the public forum could be considered the likelihood and is now influencing your original belief (prior). This likelihood also comes from a weak source of information, as anybody can share what they want in online forums without much verification. Therefore, although the likelihood provides a strong claim that aligns with the prior and could potentially strengthen your belief that aliens exist, you are still not wholly convinced and simply continue to believe that aliens probably exist (posterior). Suppose we add a different component to your initial belief that may oppose the prior:

> *You recently came across a post from a Facebook friend that explains how spy balloons are the source of several misunderstandings regarding alien ship sightings. The author of the Facebook post even states that the existence of aliens does not make logical sense for several reasons. The definite opinion of this Facebook friend contradicts what you thought, and you now consider that aliens may well be a human invention.*

In this case, you read a Facebook post refuting your aunt's story. Although the source of information is still weak (weak likelihood), it now contradicts your initial belief that aliens probably exist. The prior and likelihood in this scenario are weak and in opposition, making it hard for you to decide. Considering that the Facebook post author was more convincing than your aunt, it may change your opinion and make you think that aliens probably do not exist. This second situation is based on the same weak prior but illustrates how a weak and opposite-direction likelihood can change the posterior.

Example 2

> *You recently developed a passion for volleyball and started playing in a local team multiple days a week. While this new hobby brings you much joy, you notice that you have unintentionally lost some weight since you started. Considering that you want to maintain your current body weight, you want to determine your daily calorie needs to prevent further weight loss. You begin by watching a YouTube video from your favourite fitness influencer to know what calorie goal you should aim for and learn that most people your age only need around 2000 kcal daily. You follow this guideline and consume about 2000 kcals for the following week. By the end of that period, you notice you have lost another two pounds and decide to buy a fitness tracker (e.g., smartwatch) to determine your daily energy expenditure. After using the tracker for a week, it indicates that your average daily energy expenditure is 3000 kcals. Although you know this watch is a reliable tool to measure energy expenditure, this new calorie goal is relatively high compared to the information you got from the original YouTube video. As such, you decide to move ahead with a 2800 kcals calorie goal for the coming week to see if it helps you maintain your body weight.*

In this scenario, the information you are trying to establish is your daily calorie needs. The first source of information consulted, the YouTube video from your favorite fitness influencer, offered you a prior of 2000 kcals per day. This can be considered a weak prior because it was obtained from an unreliable and imprecise source of information, which means it has more variance. You then proceeded to collect additional data using a much more precise and reliable source of information (strong likelihood): your weekly average energy expenditure as measured by a personal fitness tracker. In this scenario, the likelihood (i.e., the data) can be considered as opposed to the prior, as it provided a much different conclusion regarding the daily calorie goal you should be using. You can also notice that, while the strong and opposite-direction likelihood has managed to change your opinion and convince you to consume a higher number of calories compared to what the prior recommended, the information from the prior was not completely erased. Instead, the prior knowledge was combined with the likelihood's information to form a new balanced conclusion. Nonetheless, because you were conscious that the information provided by the fitness tracker came from a more reliable and precise source than that of the prior, you gave more importance to the likelihood than to the prior when forming your conclusion. This explains why your final chosen daily calorie goal of 2800 kcals is closer to the likelihood (3000 kcals) than the prior (2000 kcals).

If we want to imagine a situation where a weak prior and strong likelihood would provide information that is aligned in direction, we could consider a scenario where the information given in the YouTube video would have been the same as that of the fitness tracker (i.e., 3000 kcals per day). In such a case, it would have been more likely that your conclusion regarding your daily calorie needs would be precisely 3000 kcals instead of 2800 kcals per day, meaning your initial belief would be confirmed instead of changed.

Example 3

> *You have visited the same Mediterranean country for ten days every summer for the last ten years. In those ten years, you have only witnessed two days of rain during your visits. This year you want to throw an outdoor party to celebrate your engagement, but you need to pay a considerable deposit to secure the place for the day. This decision will be affected by your belief regarding the chance that it may rain on that day. Based on the following two scenarios, we will see if you decide to pay the deposit for the day of your event or not.*

Your prior is calculated as a percentage. To do so, you would have to divide both days of rain over the total number of days of the visit. This denominator is ten days × ten years which is 100 days. Therefore, the prior is 2/100 = 2% chance of rain. Let's consider the following likelihood:

> *You arrived in the country today, just 12 hours ago, and there has not been a single drop of rain. This information makes you believe it will not rain during your engagement party.*

If you only base your decision on this data, without considering your experience, you would conclude that the likelihood of rain falling on your party's day would be 0%. However, you could also combine your ten years of knowledge (prior) with your current 12 hours of observation (likelihood), which would result in a posterior. Considering that all the rainy days you have experienced in this country can be summed up to 48 hours (two days × 24 hours) and that you have spent a total of 2412 hours there (10 days × 24 hours + the 12 hours of your current trip), your posterior for the chance of it raining on the day of your party would amount to 1.99% (48h /2412h). Since the chance of rain is so low, you decide to pay the deposit.

Another possible scenario:

> *A friend who has already arrived told you that it has constantly been raining for the last two days, the day before leaving for your trip. However, this is a friend you don't find very reliable as they tend to exaggerate everything, meaning you only trust them around 50% of the time.*

If you base your decision solely on this current information (likelihood), which would be your friend's claims, this will result in a 50% chance of rain since your friend reported rain 100% of the time, but you only trust them partially. This makes your decision to pay the deposit very difficult. However, your ten years of knowledge (inclusion of a prior) would be combined with the information from your friend to create a posterior with a different conclusion.

Since the total time of rain with the reliability considered is 72 hours (two days from your past ten years of vacation [48 hours] + two days from your friend with 50% reliability considered [24 hours]). Your full knowledge about the country is 2448 hours (ten days for ten years + two days from your friend). Your posterior for the chance of it raining on the day of your party is now 2.9% (48h + (0.5 × 48h)) /2412h). You still decide to pay the deposit as you conclude there is less than a 3% chance of rain.

Example 4

> *You are an undergraduate student with the goal of pursuing a doctorate degree and want to figure out your chances of being admitted into the graduate program of your dreams. At first, you panic because you read that only 10% of students are accepted into the program. When you hear that 85% of students in the graduate program you aim for have had a GPA on the admission of A and above, you feel very relieved. As you have a GPA above A, you consider your chances of admission to be around 85%. However, this conclusion is incorrect; it does not consider that not all students with an average above A are accepted. In fact, 5% of the students who were not accepted also had an average GPA above A.*

Indeed, 85% is the probability of having this GPA after being accepted. Still, we need to include the prior likelihood of being accepted into our equation to make an accurate prediction. Bayes theorem allows us to calculate the probabilities with a ratio of admitted students who have a GPA above A and overall students with a GPA above A (those accepted and not; see Figure 3). The numerator includes those accepted (10%) and with a GPA above A (85% of this

10%), which amounts to 10% × 85% = 8.5% of all students. For the denominator, we need to combine this number with the students who were rejected (90%) and had a GPA above A (5% of this 90%), which amounts to 4.5% of all students. As such, the final probability is 8.5%/(8.5%+4.5%) = 65.38%, meaning your actual likelihood of being accepted, considering your GPA is above A, is only 65

Example 5

A real case in the United Kingdom in the late 90s illustrates Bayesian statistics' impact in real life. Sally Clark, a mother whose two sons died of sudden infant death syndrome (SIDS), was charged with murder and sentenced to life in prison (Bertsch Mcgrayne, 2011). A pediatrician witness for the prosecution cited government statistics on the incidence of SIDS cases in a family like Sally Clark's (i.e., wealthy, non-smoking, and with a mother in the best age range) and said that this only happens in one out of 8543 cases. To calculate the chance of two children dying from SIDS, he multiplied the two statistics to arrive at one in 73 million, which was so rare that the conclusion had to be that the two children were indeed murdered (Watkins, 2000). This statistic was widely circulated in the media and strongly influenced the Conviction of Sally, who was sentenced to life in prison for murder (Hill, 2004).

However, the jury should have been able to weigh two competing explanations by comparing one highly improbable event, two siblings who died of SIDS, against another unlikely event, two siblings killed by their mother (Waite et al., 1999; Society, 2001). This way, they could have assessed whether the babies were more likely to have died of natural causes than murder. It is, therefore, essential to find the conditional probabilities of the different possible causes of death. The necessary tool, Bayes' theorem, allows the probabilities of various events to be combined to weigh their relative probabilities (Watkins, 2000).

First, regarding the risk of SIDS, the probability of a random child dying from it was 1 in 1300 then, not 1 in 8500 (Fleming et al., 2000). In addition, the probability of a second child in the same family dying becomes higher due to genetic propensity and other demographic factors, such that it increases to 1 in 100 (Guntheroth et al., 1990; Oyen et al., 1996). Concerning the probabilities of murder, according to statistics available in the United Kingdom at that time, approximately 30 children out of 650,000 annual births would have been killed by their mothers (Fleming et al., 2000). The number of double murders is estimated to be at least ten times less frequent (Hill, 2004).

If we take the hypothesis that Sally's two children died of SIDS, we realize that it is no longer a one in 73 million chance but rather a one in 130,000 chance that two children from the same family will die. As for the double murder hypothesis, the possibility that Sally's children were murdered becomes even more unlikely, at a chance of one in 218,000. However, the jury never had the opportunity to compare these two relative probabilities. Sally was released after a second appeal where evidence, not available to her defence team at trial, revealed that at the time of her second baby's death, he was suffering from a bacterial infection of the blood known to cause sudden infant death syndrome (Bertsch Mcgrayne, 2011). Sally died a few years after being released from prison, where she spent more than three years for a crime that never happened (Bertsch Mcgrayne, 2011). The grossly misleading use of statistics during her trial deprived her of the presumption of innocence.

Does the prior help or bias the results?

Suitable priors can help improve and balance situations with few or uncertain data points and, as seen in the previous examples, can lead to a different comprehension of the situation (Zyphur & Oswald, 2015; Johnson et al., 2022). Unfortunately, as many studies try to uncover new information that has not been established previously, informative and unbiased priors are not always available (Carpenter et al., 2008). There remains a level of subjectivity to choosing priors (Carpenter et al., 2008). Indeed, it is critical to use priors cautiously (Moyé, 2008), as the use of biased priors could drag bias into the conclusion. Nonetheless, solutions have been proposed to reduce this subjectivity (Depaoli et al., 2020).

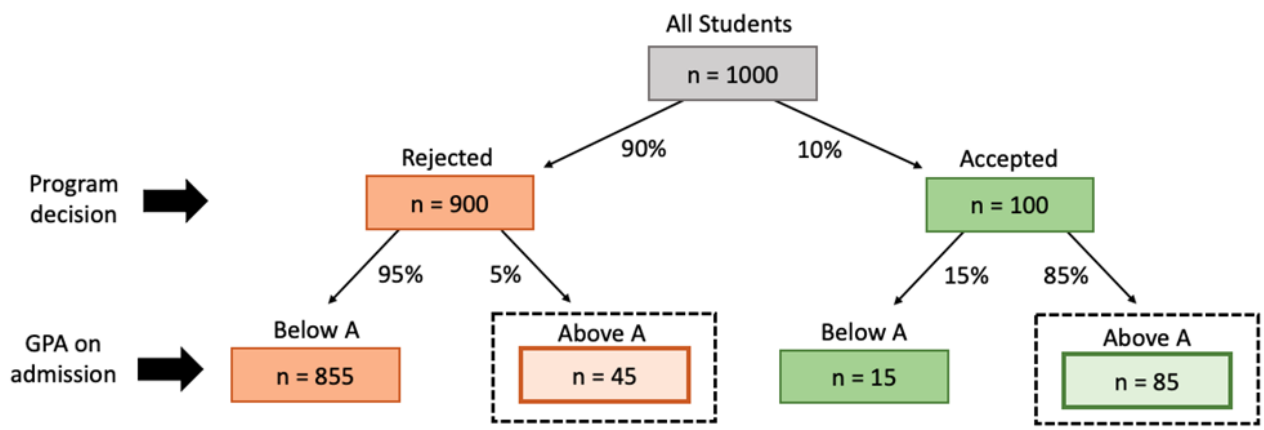
Conclusion

This paper aims to give the reader a better theoretical understanding of Bayesian reasoning and presents some core differences between the frequentist and Bayesian approaches. It highlights the relevance and potential benefits of using different statistical techniques for different situations, the importance of understanding these approaches, and their respective advantages and disadvantages.

The Bayesian approach is not taught as commonly as the frequentist approach, partly due to the limited resources available to introduce novice researchers to this alternative. This paper addresses this issue by introducing the main concepts of Bayesian reasoning through simple and intuitive examples while highlighting the importance of cumulative knowledge. These concepts include the prior, likelihood and posterior, and how their respective characteristics affect their interaction. The examples are introduced in increasing order of complexity while remaining accessible to novice learners by avoiding the use of complex statistical formulas.

Using our prior knowledge to make decisions is something humans do in their everyday life, but there would also be practical applications for using this approach in

Figure 3 ■ Visual representation of the conditional probability of admission considering GPA on admission. Using 1000 students and the probabilities described in the scenario, the likelihood of being accepted if you have a GPA on admission above A is 85/(85+45) = 85/130 = 65.38%. Probability is obtained by dividing the above A accepted students by all students with an above A GPA.



statistics. For example, in the weather scenario, the person's prior knowledge of the climate helped them make a balanced decision, even when they had a small likelihood. However, as mentioned previously, the Bayesian approach can introduce bias differently than the frequentist approach. It should, therefore, be used cautiously in situations where the priors could be biased. Nonetheless, this paper aims to facilitate the teaching and understanding of the Bayesian approach for novice researchers to help diversify the use of various statistical methods across scientific literature.

References

Bandyopadhyay, P. S. (2011). *Philosophy of statistics (handbook of philosophy of science)*. North Holland.

Beard, E., & West, R. (2017). Using Bayesian statistics in health psychology: A comment on depaoli et al. (2017). *Health psychology review, 11*(3), 298–301. doi: 10.1080/17437199.2017.1349544.

Bertsch Mcgrayne, S. (2011). *The theory that would not die: How bayes' rule cracked the enigma code, hunted down russian submarines, and emerged triumphant from two centuries of controversy*. Yale University Press.

Bland, J. M., & Altman, D. G. (1998). Statistics notes: Bayesians and frequentists. *BMJ, 317*(7166), 1151–1160. doi: 10.1136/bmj.317.7166.1151.

Bollen, K. A. (1989). Confirmatory factor analysis. *Structural equations with latent variables, 210*, 226–238.

Carpenter, J., Gajewski, B., Teel, C., & Aaronson, L. S. (2008). Bayesian data analysis: Estimating the efficacy of t'ai chi as a case study. *Nursing research, 57*(3), 214–219. doi: 10.1097/01.NNR.0000319495.59746.b8.

Cohen, J. (1994). The earth is round (p < .05). *American Psychologist, 49*(12), 997–1003. doi: 10.1037/0003-066x.49.12.997.

Cressie, N., & Wikle, C. K. (2011). *Statistics for spatio-temporal data*. Wiley.

Dancey, C. P., & Reidy, J. (2007). *Statistics without maths for psychology: Using spss for windows*. Prentice Hall.

Depaoli, S., Rus, H. M., Clifton, J. P., van de Schoot, R., & Tiemensma, J. (2017). An introduction to Bayesian statistics in health psychology. *Health psychology review, 11*(3), 248–264. doi: 10.1080/17437199.2017.1343676.

Depaoli, S., Winter, S. D., & Visser, M. (2020). The importance of prior sensitivity analysis in Bayesian statistics: Demonstrations using an interactive shiny app. *Frontiers in psychology, 11*, 608045. doi: 10.3389/fpsyg.2020.608045.

Dogucu, M., & Hu, J. (2022). The current state of undergraduate Bayesian education and recommendations for the future. *The American Statistician, 74*, 405–413. doi: 10.1080/00031305.2022.2089232.

Epstein, L. G., & Schneider, M. (2007). Learning under ambiguity. *The Review of Economic Studies, 74*(4), 1275–1303.

Faber, J., & Fonseca, L. M. (2014). How sample size influences research outcomes. *Evidence-based Orthodontics, 19*(4), 27–29. doi: 10.1590/2176-9451.19.4.027-029.ebo.

Fleming, P., Bacon, C., Blair, P., & Berry, P. (2000). *Sudden unexpected deaths in infancy, the cesdi studies 1993-1996*. The Stationary Office.

Guntheroth, W. G., Lohmann, R., & Spiers, P. S. (1990). Risk of sudden infant death syndrome in subsequent siblings. *The Journal of Pediatrics*, *116*(4), 520–524. doi: 10.1016/s0022-3476(05)81596-3.

Hackenberger, B. K. (2019). Bayes or not bayes, is this the question? *Croatian medical journal*, *60*(1), 50–52. doi: 10.3325/cmj.2019.60.50.

Hill, R. (2004). Multiple sudden infant deaths – coincidence or beyond coincidence? *Paediatric and Perinatal Epidemiology*, *18*(5), 320–326. doi: 10.1111/j.1365-3016.2004.00560.x.

Jaynes, E. T. (1976). Confidence intervals vs Bayesian intervals. In W. L. Harper & C. A. Hooker (Eds.), *Foundations of probability theory, statistical inference, and statistical theories of science, vol* (pp. 175–257). Springer.

Johnson, A. A., Ott, M. Q., & Dogucu, M. (2022). *Bayes rules!: An introduction to applied Bayesian modeling (1st ed.)* Chapman; Hall/CRC. doi: 10.1201/9780429288340.

Kaplan, D. (2014). *Bayesian statistics for the social sciences*. The Guilford Press.

Kurt, W. (2019). *Bayesian statistics the fun way: Understanding statistics and probability with star wars, lego, and rubber ducks*. No Starch Press.

Lecoutre, B. (2006). Training students and researchers in Bayesian methods. *Journal of Data Science*, *4*(2), 207–232. doi: 10.6339/jds.2006.04(2).246.

Lecoutre, M.-P., Poitevineau, J., & Lecoutre, B. (2003). Even statisticians are not immune to misinterpretations of null hypothesis significance tests. *International Journal of Psychology*, *38*, 37–45. doi: 10.1080/00207590244000250.

Lele, S. R. (2020). How should we quantify uncertainty in statistical inference? *Frontiers in Ecology and Evolution*, *8*, 1–99. doi: 10.3389/fevo.2020.00035.

MacKinnon, D. P., Fairchild, A. J., & Fritz, M. S. (2007). Mediation analysis. *Annual review of psychology*, *58*, 593–614. doi: 10.1146/annurev.psych.58.110405.085542.

Moyé, L. A. (2008). Bayesians in clinical trials: Asleep at the switch. *Statistics in medicine*, *27*(4), 469–489. doi: 10.1002/sim.2928.

Oyen, N., Skjaerven, R., & Irgens, L. M. (1996). Population-based recurrence risk of sudden infant death syndrome compared with other infant and fetal deaths. *American Journal of Epidemiology*, *144*(3), 300–305. doi: 10.1093/oxfordjournals.aje.a008925.

Reimherr, M. L., Meng, X., & Nicolae, D. L. (2014). Assessing prior informativeness; prior likelihood conflict.

Society, T. R. S. (2001). *Royal statistical society concerned by issues raised in sally clark case*. https://web.archive.org/web/20110824151124/http://www.rss.org.uk/uploadedfiles/documentlibrary/744.pdf

Tukey, J. W. (1991). The philosophy of multiple comparisons. *Statistical Science*, *6*(1), 100–116. doi: 10.1214/ss/1177011945.

van de Schoot, R., & Depaoli, S. (2014). Bayesian analyses: Where to start and what to report. *The European Health Psychologist*, *17*(2), 75–84.

van de Schoot, R., Winter, S. D., Ryan, O., Zondervan-Zwijnenburg, M., & Depaoli, S. (2017). A systematic review of Bayesian articles in psychology: The last 25 years. *Psychological Methods*, *22*(2), 217–239. doi: 10.1037/met0000100.

Waite, A. J., Emery, J. L., McKenzie, A., Carpenter, R. G., Coombs, R. C., & Daman-Willems, C. (1999). Repeat cot deaths in families enrolled on the coni programme 1989-1997. *Pediatric Research*, *45*, 12A–12A. doi: 10.1203/00006450-199905020-00046.

Watkins, S. J. (2000). Conviction by mathematical error? *BMJ*, *320*(7226), 2–3. doi: 10.1136/bmj.320.7226.2.

Winkler, R. L. (2001). Why Bayesian analysis hasn't caught on in healthcare decision making. *International journal of technology assessment in health care*, *17*(1), 56–66. doi: 10.1017/s026646230110406x.

Zyphur, M. J., & Oswald, F. L. (2015). Bayesian estimation and inference: A user's guide. *Journal of Management*, *41*(2), 390–420. doi: 10.1177/0149206313501200.

Citation

Pétrin-Pomerleau, P., Vincent, C., Garcia Mairena, P. M., Yilmaz, E., Théberge Charbonneau, A., Husereau, T., Niyonkuru, G., & Jacob, G. (2023). The bayesian approach is intuitive conditionally to prior exposition to these examples. *The Quantitative Methods for Psychology*, *19*(3), 244–253. doi: 10.20982/tqmp.19.3.p244.

Received: 20/07/2023 ~ Accepted: 04/08/2024