

BAYESIAN ANALYSIS AS A FRAMEWORK FOR (LEGAL) THINKING

Neil Chilson*

CITE AS: 1 GEO. L. TECH. REV. 89 (2016)

<https://perma.cc/926E-U882>

INTRODUCTION.....	89
THE BAYESIAN FRAMEWORK.....	90
CRITICISMS OF THE BAYESIAN APPROACH.....	93
APPLYING BAYESIAN ANALYSIS.....	94

INTRODUCTION

Every day, lawyers, regulators, and policy advocates must use information about a situation to evaluate, predict, and draw conclusions about the world around them. The problem is ubiquitous: You have a hypothesis that attempts to explain something in the real world – perhaps the cause of a disease, the guilt of an accused criminal, or the effect of a policy or legislative decision. You also possess some data about the issue at hand. Does this data support or undermine the hypothesis? This is the process of inference. Each of us makes thousands of inferences each day. For example, even without hearing the morning weather report, we might reasonably choose to wear a sweater in February but select a t-shirt in August. We infer the weather conditions based on our hypothesis about the local weather. For such low-stakes decisions, we do not usually formally test our hypotheses. However, when the stakes are high, or when there are many data points that don't clearly point in one direction, we can often improve our inferences by applying the statistical tools of Bayesian analysis.

Bayesian analysis is one of two toolsets statisticians developed to test hypotheses. Most readers will have encountered the frequentist approach to hypothesis testing, although they may not know it by that name. However, non-experts may not know about the increasingly popular second toolset: Bayesian

* Attorney Advisor to Commissioner Maureen K. Ohlhausen at the Federal Trade Commission; J.D., The George Washington University Law School, 2007; M.S. in Computer Science, The University of Illinois, Urbana-Champaign, 2005; B.S. in Computer Science, Harding University, 1999. The views expressed here are those of the author and should not be attributed to Commissioner Ohlhausen or to the Federal Trade Commission. © 2016, Neil Chilson.

analysis.¹ Although Bayesian analysis was anathema to professional statisticians for more than a century, today its applications are ubiquitous.² This is because Bayesian analysis offers a practical way to transform large volumes of data into actionable recommendations. Machine learning, neural networks, and pattern recognition algorithms are built on Bayesian principles.³ These and other Bayesian tools form the computational underpinnings of common applications such as spam filters, movie recommendation engines, cancer diagnosis, language translation, codebreaking, and self-driving automobiles.

Understanding the basic principles of Bayesian analysis will help the modern lawyer grapple with such pervasive technologies. Bayesian analysis also could enhance legal and policy decisions and improve critical thinking skills generally.

THE BAYESIAN FRAMEWORK

At the core of Bayesian analysis is Bayes' Theorem, named after Reverend Thomas Bayes, who first discovered it in the 18th century. Bayes' Theorem formalizes a rather common-sense procedure: when we gather new data about a situation, we use it to update our existing belief, creating an improved belief about that situation. But while this sounds like common sense, Bayesian analysis can be at its most useful when it produces counterintuitive results.

We can unpack three important characteristics of Bayesian analysis through an example. Imagine that while you are reading this article in your home or office, the fire alarm suddenly starts sounding. Given this new information about the state of the world, estimate the probability (90%? 50%? 2%?) that there is a dangerous fire in your building.

¹ The roots of today's frequentist and Bayesian statistical tools date to the mid-1700s. (Reverend Thomas Bayes, namesake of the Bayesian approach, discovered the core rule in the 1740s.) Due to fundamental philosophical differences between the two approaches, statisticians have generally considered themselves either frequentists or Bayesian, with frequentists overwhelmingly dominating the field until recent decades. This article will barely touch on the philosophical differences between the camps and will not at all address its fascinating history. Excellent accounts are available, including Sharon Bertsch McGrayne's *The Theory that Would Not Die: How Bayes' Rule Cracked the Enigma Code, Hunted Down Russian Submarines & Emerged Triumphant from Two Centuries of Controversy*.

² See, e.g., F.D. Flam, *The Odds, Continually Updated*, N.Y. TIMES (Sept. 29, 2014), <http://www.nytimes.com/2014/09/30/science/the-odds-continually-updated.html>.

³ See generally SHARON BERTSCH MCGRAYNE, *THE THEORY THAT WOULDN'T DIE* 233–52 (2011).

How did you choose that percentage? You probably came to it quite intuitively, but let's examine the specific categories of information that may have informed that intuition. First, you have an existing belief, a *prior probability*, about the state of the world: absent an alarm, you know that dangerous fires are rather unlikely as a general matter. Second, you know that there are multiple situations other than a fire that might trigger a fire alarm, and that some of these triggers are more probable than others. There might be a dangerous fire, in which case the alarm is a *true positive*: it indicates there is a fire, and indeed there is. But perhaps building administrators are testing the alarm. Perhaps someone burned bacon. Perhaps someone pulled the fire alarm as a prank. For each of these causes the alarm is a *false positive*: it is indicating the existence of a fire, but there is no fire. You consider the likelihood of these true and false positives in light of your prior belief about the state of the world, and come to a new belief, a *posterior probability*, about the state of the world. Having derived this inference about the world, you probably aren't done: you will continue to gather additional information—Is there smoke? Is there an announcement? Do you hear fire engines?—and use this new information to update your belief about the state of the world around you.

This example demonstrates the three key characteristics of Bayesian analysis. First, when applying Bayesian analysis to new evidence, our prior beliefs – how likely are building fires? – are an input. They affect how new information changes our conclusion. Second, Bayesian analysis weighs the likelihood of each potential explanation for the new evidence, some of which conflict with and some of which support our prior beliefs. Third, and finally, Bayesian analysis is iterative, meaning we continually update our conclusions by taking our previous analysis's output and using it as input for the next round of analysis.

Each of these three characteristics is reflected in Bayes' Theorem:

$$P(A|B) = \frac{P(B|A) \cdot P(A)}{P(B)}$$

where $P(A/B)$ means the probability of A given B.⁴

Let's walk through our fire alarm example to concretize the abstract symbols. We are estimating the probability that there is a dangerous fire given that we hear a fire alarm. Thus, referring to the left side of the formula above, we are estimating $P(A/B)$, or the probability of A (a fire) given B (a sounding

⁴ This is a simplified version of the equation. The more explicit version expands the denominator, $P(B)$, into the mathematically equivalent $P(B/A) \cdot P(A) + P(B/\sim A) \cdot P(\sim A)$.

fire alarm). The right side of the equation shows how to find $P(A/B)$. It has three parts. First is $P(A)$, the *prior probability* that stands for how probable are building fires as a general matter. As mentioned before, the probability is likely quite low. Erring on the side of caution, let us assume that 1% of buildings are on fire at any one time. Second is $P(B)$, which in our example is the probability of a fire alarm going off. This includes the probability of the fire alarm going off because there is a fire (a *true positive*) plus the probability of the alarm going off for any other reason (*false positives*). Based on experience, I would guess that most fire alarms are false positives, so we would expect this probability to be quite a bit larger than the 1% percent of buildings on fire. Even so, fire alarms aren't that common. Let's say that 5% of buildings have an active fire alarm at any time. The third component is $P(B/A)$, which in our example is the probability of an alarm going off assuming there is a fire in the building. One hopes that this is quite high: while fire alarms may go off for many other reasons, when a fire is burning we want an alarm that is certain to go off. Let's guess that fire alarms activate to 99% of building fires.

Inputting these values into the equation, we get a $(.99 * .01)/.05 = .198 = 19.8\%$ probability, or approximately a one in five chance that the building is burning given that a fire alarm is sounding. Your estimate may be higher or lower depending on your estimates for each value. However, if, like me, you estimated a relatively low likelihood of a fire given the alarm, that is probably because your experiences with fire alarms (like mine) suggest that false positives are quite common – fire alarms go off for many reasons other than actual fires.

In this example, Bayes appears to have simply confirmed the common-sense conclusion that fire alarms often don't mean the building is burning. So why is Bayesian analysis useful if it is essentially common sense? First, this fire alarm example is very straightforward, with only a single piece of new data (the fire alarm is sounding). Bayesian analysis can be used for much more complex problems involving many new data points with thousand or millions of possible values where there is no obvious intuitive answer. Our intuitions prove less useful with such complex data sets.

Second, while we all have experience with false positives in fire alarms, our intuitions can lead us astray in areas where we have less experience with false positives. Bayesian analysis can help overcome such incorrect intuitions. One of the best examples comes from a controversial 2009 U.S. Preventative Services Task Force recommendation.⁵ The Task Force recommended *against*

⁵ *Final Update Summary: Breast Cancer: Screening*, U.S. PREVENTATIVE SERVS. TASK FORCE (Jan. 2016),

routine biennial mammograms for women between 40 and 50 with no other risk factors. It recommended this even though mammograms are relatively accurate: such tests properly identify approximately 75% of women with breast cancer, although they return false positives (misidentify cancer in women who are cancer-free) about 10% of the time. Despite the apparent accuracy of the tests, the task force found that a woman under 50 without additional risk factors who had a positive mammogram test result was highly *unlikely* to have breast cancer.⁶

This counterintuitive result can be made more intuitive by focusing on the false positive component of the Bayesian formula. Like burning buildings in our fire alarm example, women under 50 with breast cancer are very rare – approximately 40 out of 10,000 – meaning 9,960 out of 10,000 women don't have breast cancer. But if we tested all 10,000 women, a 10% false positive rate would produce approximately 996 false positives – dwarfing the number of women with cancer. Like our fire alarm example, when testing an extremely rare condition, even relatively accurate tests produce many false positives.

The Task Force's report explained that these false positives were a major cost of unnecessary screening, imposing “psychological harms, additional medical visits, imaging, and biopsies in women without cancer, inconvenience due to false-positive screening results, harms of unnecessary treatment, and radiation exposure.”⁷ The counterintuitive recommendation against such routine testing was controversial, but would not have been if more people understood Bayes' Theorem and how our intuitions can fail to account for false positives.

CRITICISMS OF THE BAYESIAN APPROACH

Bayesian analysis has faced strong criticism for centuries. Indeed, the approach was practically taboo among professional statisticians for much of its history, even though non-statistician practitioners periodically used it to solve real-world problems.⁸ Statisticians in the long-dominant frequentist school of

<https://www.uspreventiveservicestaskforce.org/Page/Document/UpdateSummaryFinal/breast-cancer-screening>.

⁶ This example draws on the characterization of this research in McGrayne, *supra* note 3, at 259.

⁷ *Screening for Breast Cancer Using Film Mammography: Clinical Summary of 2009 U.S. Preventive Services Task Force Recommendation*, U.S. PREVENTATIVE SERVS. TASK FORCE (2009), <https://www.uspreventiveservicestaskforce.org/Home/GetFileByID/248>.

⁸ Historical applications include breaking the Nazi's Enigma code and identifying the causes of lung cancer and heart disease. *See, e.g.*, McGrayne, *supra* note 3.

probability and hypothesis testing acknowledge the mathematical correctness of Bayes' Theorem. However, they raised several concerns about the usefulness of Bayesian analysis. Much of this opposition was a practical matter. The above examples are relatively simple applications of Bayes' Theorem, but for more complex problems involving multiple new data points with many possible values, Bayesian analysis requires substantial computing power not available until the modern computer age. Traditional frequentist tools were simply more practical prior to computers.

More fundamentally, frequentists were highly skeptical of Bayesian reliance on prior probabilities, which they argued are overly subjective and thus fail to produce objective, scientific results. Bayesians counter that all analysis relies on prior knowledge and experience, and the Bayesian approach exposes such reliance rather than conceal it.

APPLYING BAYESIAN ANALYSIS

Engineers have enthusiastically applied Bayesian analysis to build new and useful tools, some of which we have already mentioned. But as computation has become a useful problem-solving tool across all areas of human inquiry, Bayesian analysis has broadening implications for law and policy. The Bayesian approach promises important legal and policy applications and adopting a Bayesian mindset can strengthen critical thinking skills generally.

Criminal or Civil Trials. Bayesian analysis has significant potential – largely unrealized – to assist courts in determining civil or criminal liability.⁹ Some have argued that it can be helpful in assessing the overall significance of an accumulation of small pieces of evidence.¹⁰ Others have suggested a Bayesian approach to assessing expert testimony.¹¹ In the 1970s, academics heatedly debated the use of Bayesian methods in courtrooms.¹² Thanks to the

⁹ Michael O. Finkelstein & William B. Fairley, *A Bayesian Approach to Identification Evidence*, 83 HARV. L. REV. 489 (1970); *see also*, Kristy L. Fields, *Toward A Bayesian Analysis of Recanted Eyewitness Identification Testimony*, 88 N.Y.U. L. REV. 1769 (2013) (arguing that “Bayesian analysis can, and should, be used to evaluate [uncertain eyewitness testimony],” as “use of an objective method of analysis can ameliorate cognitive biases and implicit mistrust of recantation evidence.”).

¹⁰ *Id.*; *see also*, Jason R. Bent, *Hidden Priors: Toward A Unifying Theory of Systemic Disparate Treatment Law*, 91 DENV. U. L. REV. 807 (2014) (enumerating a “coherent theory of systemic disparate treatment that embraces Bayesian priors.”).

¹¹ Bruce Abramson, *Blue Smoke or Science? The Challenge of Assessing Expertise Offered as Advocacy*, 22 WHITTIER L. REV. 723 (2001).

¹² *See, e.g.*, MICHAEL O. FINKELSTEIN & BRUCE LEVIN, *STATISTICS FOR LAWYERS* 84–5 (3rd ed. 2015).

information revolution, the practical applications of Bayesian techniques are far more powerful than in the 1970s,¹³ so it may be time to revive that debate.

Courts present hurdles to the use of Bayesian analysis. Statistics of all kinds have often failed to persuade juries, in part due to the lack of statistical understanding by judges, jurors, and even by lawyers attempting to use them.¹⁴ However, as Bayesian-based technologies such as machine learning become more widely applied, including in the context of law enforcement,¹⁵ we will surely see court challenges that involve the technology.

Policy Analysis. Bayesian analysis can help inform policy decisions, as demonstrated by the earlier discussion of breast cancer screening recommendations. Bayesian analysis is particularly useful in performing meta-analysis of many different studies to identify trends in research. It can also be helpful in estimating the probability of one-time events (such as an accidental nuclear detonation or catastrophic climate change), analyzing phenomena where data is sparse, or interpreting experiments that are not easily reproduced.¹⁶ In these areas, the more traditional statistical tools of frequentism falter. Indeed, frequentists have argued that it makes no sense to talk about the “probability” of an event that has never occurred because we lack any information about the long-term frequency of such an event.¹⁷ Yet many important policy decisions involve analyzing a wide array of different but related research, or evaluating unique circumstances with little hard data. Bayesian analysis cannot provide simple answers for such complex problems, but it can provide a framework for thinking through the issues involved.¹⁸

¹³ See generally, Maggie Wittlin, *Hindsight Evidence*, 116 COLUM. L. REV. 1323 (2016) (discussing the applicability of Bayesian reasoning to decision-making by the members of a jury); Ian Ayres & Barry Nalebuff, *The Rule of Probabilities: A Practical Approach for Applying Bayes' Rule to the Analysis of DNA Evidence*, 67 STAN. L. REV. 1447 (2015) (discussing the benefit of applying Bayes' Rule to the use of probabilistic DNA matching).

¹⁴ See generally Fenton et al., *Bayes and the Law*, 3 ANN. REV. OF STAT. AND ITS APPLICATION 51–77 (June 2016).

¹⁵ See Andrew Guthrie Ferguson, *Big Data and Predictive Reasonable Suspicion*, 163 U. PA. L. REV. 327 (2015).

¹⁶ See, e.g., Fred Charles Iklé, Gerald J. Aronson, & Albert Madansky, *On the Risk of Accidental or Unauthorized Nuclear Detonation*, THE RAND CORP. WORKING PAPER (Oct. 15, 1958),

http://www.rand.org/content/dam/rand/pubs/research_memoranda/2006/RM2251.pdf. This paper is discussed at length at McGrayne, *supra* note 3, at 119 *et seq.*

¹⁷ McGrayne *supra* note 3, at 119–121.

¹⁸ See, e.g., Jason R. Bent, *P-Values, Priors, and Procedure in Antidiscrimination Law*, 63 BUFF. L. REV. 85, 89 (2015) (arguing that the “time [has] finally come for a Bayesian revolution in employment discrimination law.”).

Case Selection. Regulators and law enforcement agencies could benefit from using Bayesian analysis as a case-selection tool, at least in certain circumstances. For example, the Federal Trade Commission evaluates thousands of potentially false or deceptive advertisements each year – more than it has the resources to pursue. Bayesian analysis could help make case selection more intellectually rigorous. Advertising claims fall on a continuum from fraudulent to truthful and informative. Similarly, companies base their advertising claims on evidence that ranges from non-existent, to mixed, to conclusive – either in support of or against the claim. A Bayesian approach to case selection would take into account that certain types of common claims – rapid weight loss, for example – are nearly always misleading, at best. Thus, there would be a strong prior probability of a violation that could only be outweighed by powerful evidence substantiating the claim. Other claims, such as mild weight loss benefits, often have mixed evidence, and therefore the prior probability of a violation would be weaker. Bayesian analysis could help most in areas where the company offers several partially flawed studies that generally support their claims. These are the areas where traditional informal case selection methods might be most improved.¹⁹

In the private sector, firm lawyers and in-house counsel could use a similar approach to evaluate the risk of legal liability under uncertain conditions. Such assessments could both inform compliance advice to clients and shape litigation and settlement strategies.

Critical Thinking. Even forgoing the formal mathematics of Bayes' Theorem, the legal career of any lawyer could benefit from embracing a Bayesian mindset when considering evidence, and making decisions. There are three lessons from Bayes' that can benefit anyone who needs to evaluate evidence and make decisions. **First**, carefully evaluate the impact of prior beliefs when reviewing evidence. Bayes' Theorem requires that we make explicit what is often implicit: we interpret evidence in light of our own prior beliefs. This “bias” may be appropriate, as our prior experiences can be very

¹⁹ See, e.g., Stephen Charest, *Bayesian Approaches to the Precautionary Principle*, 12 DUKE ENVTL. L. & POL'Y F. 265, 270 (2002) (arguing that “that a Bayesian approach is not only the appropriate method to assess truly uncertain risks,” but also “is uniquely suited to promoting intellectual due process principles” in regulatory decision-making); Geoffrey Christopher Rapp, *Intelligence Design: An Analysis of the SEC's New Office of Market Intelligence and Its Goal of Using Big Data to Improve Securities Enforcement*, 82 U. CIN. L. REV. 415, 422 (2013) (discussing the use of Bayesian techniques in intelligence and securities enforcement).

informative. However, becoming more aware of our own inherent bias can help us recognize when that bias may be inappropriate or even misleading.

Second, thoroughly consider other explanations for evidence. Bayesian analysis requires researchers to evaluate the likelihood that the evidence might be explained by something other than their preferred hypothesis. It requires understanding other potential explanations, such as false positives. Simply stepping back to consider other potential explanations of the same data can prevent myopic decision-making.

Third, and finally, think of decision-making as an iterative process that is never finished. At its core, Bayesian analysis relies on repeatedly integrating new evidence with what one already knows. New evidence can always help refine a prediction, and no prediction is ever perfected. Bayesian analysis suggests that we can absorb new evidence in rigorous and principled ways while recognizing that 100% certainty is rarely, if ever, warranted. Instead, policymaking and law are part of life's constant journey toward a better, but never perfect, understanding of the world. Bayesian analysis can help guide us on that journey.