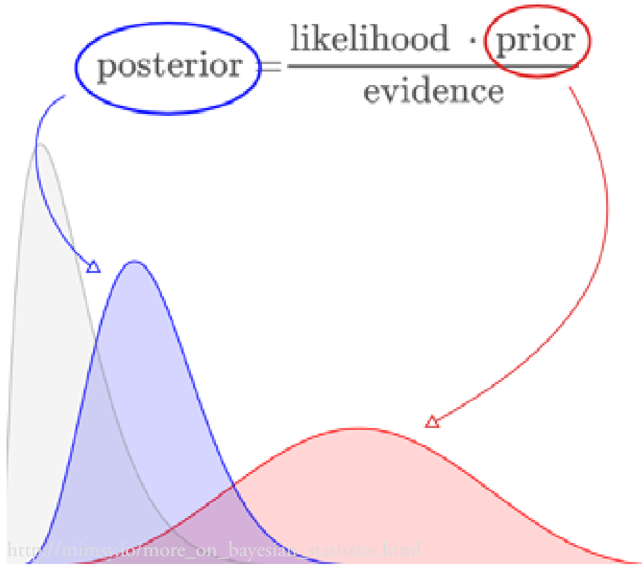
Five-minute Science Lesson:

Bayesian Analysis and Scientific Credibility Assessment Testing



by Raymond Nelson

Scientific tests are used to quantify any phenomena that cannot be subject to perfect deterministic observation or direct physical measurement. Because deception and truth-telling cannot be measured directly, and because any ability to observe deception or truth deterministically would obviate the need for testing, scientific credibility assessment tests are an example of the value and need for scientific tests to assist with intelligent decision making under uncertain conditions. Scientific tests are not expected to be infallible and are instead fundamentally dependent upon statistics and probabilistic inference. Statistical and probabilistic inference can be thought of as existing in two important discussions: frequentist inference, and Bayesian inference.

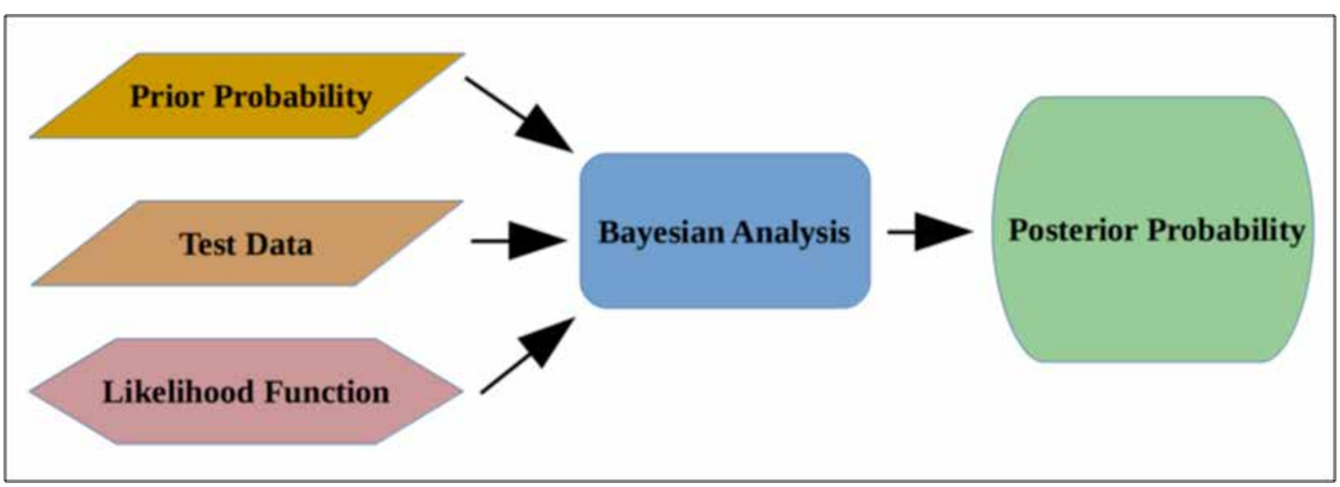
Frequentist inference – concerned with counting the frequency of repeatable events, and with errors of measurement when evaluating the universe of reality that is assumed to exist in one fixed way – has formed the basis of the scientific method of hypothesis testing for the greater part of the last century. Although useful for the purpose of understanding measurement error, frequentist inference has proved to be a confusing basis for classification and decision-making.

Bayesian inference – the basis of Bayesian analysis was developed out of the need to make intelligent and quantified decisions under uncertain conditions that are un-repeatable or unobservable, such as the most likely location for a lost item, or the outcome of an election or athletic tournament. Unlike frequentist analysis, Bayesian inference can also be used to make inferences about *the likely cause* of some observed data. Bayesian inference therefore has great practical value in all areas of science and decision-making under uncertain conditions.

Bayesian analysis uses both evidence, in the form of test data, and probability statements to make an evidence-based estimate about something that is unknown or unobservable. In Bayesian analysis, this is referred to as an *unknown parameter,* such as a class or category. Bayesian analysis is used in many scientific and statistical decision-making contexts. For example, Bayesian analysis can be used to filter email messages, to discriminate between *spam* (junk email) and *ham* (useful email) when it is unknown how an email recipient will actually classify a particular message. In the credibility assessment context, Bayesian analysis can be used to calculate probability estimates for deception or truth-telling when it is unknown whether a person is actually truthful or deceptive.

Bayesian analysis uses three input values: 1) a prior probability, 2) the test data, and 3) a likelihood function. The posterior probability is a statistical description of the unknown parameter of interest, such as the posterior probability of deception or truth-telling. Figure 1 shows a flow chart of the inputs and output of a Bayesian analysis.

Figure 1. Bayesian analysis



The *prior probability* is a statistical or probability description of our knowledge of the parameter of interest prior to completing the test and Bayesian analysis. Good information about the prior probability is sometimes available, though it often occurs that very little information is known to support one class or conclusion over another. The prior probability can be estimated from the number of possible outcome classes when no prior information, or very little good prior information, is available. For example: when little prior information is available or known about whether an individual is actually deceptive or truthful, the prior probability can be estimated from the possible conclusions: deception or truth-telling – for which the prior probabilities are .5 and .5.[1,2] The purpose of Bayesian analysis is to update this prior probability.

Test data used as input for the Bayesian analysis are after feature extraction and after reduction and aggregation of the data to numerical values that can have more practical value than raw data. In the polygraph testing context this means that input test data are the numerical scores for the grand total and question subtotals.

A *likelihood function* is a mathematical device – often in the form of a formula, algorithm, or reference table– used to calculate a numerical or statistical value for the test data. For example, ESS-M reference tables (Nelson, 2017) are a form of theoretical reference distribution – calculated using facts and information subject to mathematical and proof, under the analytic theory of the polygraph test[3,4]. In contrast, the likelihood function for ESS scores of polygraph test data is a set of normative reference tables (Nelson, et al., 2011), calculated from empirical sampling data consisting of confirmed

1 These prior probabilities can also be referred to as a *prior probability distribution* because the values .5 and .5 represent the distribution of all of the possible classes.

2 These probabilities can also be transformed mathematically so they can be expressed as the *odds*. This is done by taking each of the probabilities and dividing by its mathematical compliment (e.g., odds = .5 / (1 - .5) = 1). When done this way, the odds are expressed in relation to the value 1 (e.g., 1 to 1). Also, odds can be transformed mathematically to probabilities by dividing the odds by odds + 1. For example: probability = 1 / (1 + 1) = .5.

3 The basic analytic theory of the polygraph test holds that greater changes in physiological activity are loaded at different types of test stimuli as function of deception or truth-telling in response to the relevant or investigation target stimuli. Refer to Nelson (2016) for a more complete discussion of this analytic theory of polygraph testing.

4 ESS scores are three-position (+ 0 or –) non-parametric values that are assigned according to the differential loading of changes in physiological activity in response to relevant and comparison stimuli, using multiple recording sensors and multiple iterations of a stimulus question sequence that includes multiple versions of the test questions. The resulting ESS-M distribution is a multinomial distribution and can be used as a likelihood function for Bayesian analysis with any population or group for which the basic theory of the polygraph can be assumed to hold true.

guilty and confirmed innocent field cases[5].

Bayesian analysis involves the use of *Bayes' theorem*. Bayes theorem is among the most practical and useful of all mathematical and statistical formulae. It is used to update or improve the effectiveness of conclusions, predications, or classifications compared to those that could be made using only the prior probability distribution. An important aspect of Bayesian analysis is that the notion of *probability* is taken to mean *degree of belief* in an idea or conclusion. This is in contrast to the frequentist notion of probability which refers to a *proportion of observed outcomes* – which limits the discussion of probabilities to things that can be observed and repeated[6].

The output or result of Bayesian analysis is a *posterior probability.* This can also be thought of as a *posterior probability distribution* when considering all posterior probabilities for all possible classes. A convenient aspect of Bayesian statistical analysis is that statistical or probabilistic results from Bayesian analysis can be intuitively useful because they can provide a direct and convenient estimate of the effect size of interest, for example: the *odds of deception* or *odds of truth-telling*.

Experts who work with and use scientific credibility assessment (lie detection) tests, and the results from these tests, can be expected to have some reasonable foundation for understanding the basics of Bayesian analysis and statistical decision-making under uncertain conditions. It is hoped that this short description will be of some value to field practitioners who wish to increase their fluency with these concepts.

---

5 Empirical normative data can be intuitively useful because information is obtained from actual tests with actual persons. Empirical norms are of unknown representativeness when used with persons outside the normative sampling group. Also, sampling distributions are always an imperfect representation of the population distribution, though we can rely on the *law of large numbers* and the use of numerous sampling distributions to converge towards the population group of interest.

6 Frequentist statistical methods cannot be used to assign a probability to events that cannot be repeated, such as sports outcomes or election outcomes. Also, frequentist statistical methods cannot be used to assign probabilities to ideas such as hypotheses or class decisions. For this reason, Bayesian analytics has important practical utility because it can be used to produce reproducible probability estimates for many practical and interesting things that cannot be observed and counted, such as the most likelihood of deception or truth-telling as the cause for some observed data during polygraphic credibility assessment testing.

References

Bayes, T. & Price, R. (1763). An essay towards solving a problem in the doctrine of chances. *Philosophical Transactions of the Royal Society of London, 53,* 370–418.

Berger, J. O. (1985). *Statistical Decision Theory and Bayesian Analysis.* Springer.

Bernardo, J. & Smith, A. (1994). *Bayesian Theory.* John Wiley & Sons.

Casella, G. (1985). An introduction to empirical Bayes data analysis. *American Statistician, 39 (2)*, 83–87.

Downey, A. B. (2012). *Think Bayes. Bayesian Statistics Made Simple.* Green Tea Press.

Efron, B. (1986). Why isn't everyone a Bayesian? *American Statistician. 40(1).* 1-5.

Gelman, A., Carlin, J. B., Stern, H. S., Dunson, D. B., Vehtary, A., & Rubin, D. B. (2014). *Bayesian Data Analysis.* CRC Press.

Gill, J. (2007). *Bayesian Methods: A Social and Behavioral Sciences Approach. Second Edition.* Chapman and Hall/CRC.

Laplace, S. P. (1812). Théorie analytique des probabilités. Paris: Courier.

Nelson, R. (2016). Scientific (Analytic) Theory of Polygraph Testing. *APA Magazine, 49(5)*, 69-82.

Nelson, R. (2017). Multinomial reference distributions for the Empirical Scoring System. *Polygraph & Forensic Credibility Assessment, 46 (2),* 81-115.

Nelson, R. Handler, M. Shaw, P., Gougler, M., Blalock, B., Russell, C., Cushman, B., & Oelrich, M. (2011). Using the Empirical Scoring System, *Polygraph, 40(2)*.

Pericchi, L. & Walley, P. (1991). Robust bayesian credible intervals and prior ignorance. *International Statistics Review 58(1)*, 1-23.

Stone, J. (2013). *Bayes' Rule: A Tutorial Introduction to Bayesian Analysis.* Sebtel Press.

Winkler, R. L. (1972). *An Introduction to Bayesian Inference and Decision.* Holt McDougal.