

Bayesian Concepts: An Introduction

John Amrhein and Fei Wang, McDougall Scientific Ltd.

ABSTRACT

You employ Bayesian concepts to navigate your everyday life, perhaps without being aware that you are doing so. You rely on past experiences to assess risk, assign probable cause, navigate uncertainty, and predict the future. Yet, as a statistician, economist, epidemiologist, or data scientist, you hold tight to your frequentist methods. Why? This paper explores the philosophy of Bayesian reasoning, explains advantages to applying Bayes' rule, and confronts the criticism of subjective Bayesian priors.

INTRODUCTION

This paper is concerned with using statistics for decision support. Specifically, when one is faced with making a decision on a course of action, he/she assesses the probabilities of possible outcomes when there is uncertainty about those outcomes. In this context, statistics is reliant on probabilities. Within the profession of statistics, there are two schools of thought regarding the definition and interpretation of probabilities; frequentist, and Bayesian.

Bayesian reasoning combines past experience with current information to assign probable cause and assess risk of an (un)wanted effect. For example, you may know a street in your town where you can park your car without paying for a permit because past experience indicates you will not get caught. Elsewhere, you pay for the permit because the risk is too high that you will get caught and have to pay a fine. In either case you combine your experiential knowledge with the known state of current affairs to decide whether to pay for a temporary permit. We process information in a similar manner to conclude probable cause for an illness or injury by considering past experience and gathering information from subject matter experts and peers. Bayes' Theorem is a mathematical construct that reflects this manner of processing information and making decisions.

Bayesian statistical methods are often described by how they differ from frequentist methods. There are a few fundamental differences:

	Frequentist	Bayesian
Population parameters	Fixed but unknown	Random
Experimental data	Random yet repeatable	Fixed
Inferences	$P\{\mathbf{X} H_0(\theta)\}$	$P\{\theta \mathbf{X}\}$
Interpretation	Mathematical frequency: Over repeated sampling, <u>an</u> estimated interval will capture the mean, say, 95 out of 100 times.	Probabilistic belief: The probability that <u>the</u> parameter's value is within the interval is, say, 0.95.
Prior information	Ignored or used indirectly (e.g. estimating sample sizes)	Directly incorporated via $P\{\theta\}$

Frequentist methods might have their highest value in tightly controlled experimental situations (rolling dice), when repeatability is possible, even probable. However, in situations in which repeatability is questionable, for example in biological studies, Bayesian methods are advantageous because data are treated as fixed and repeatability is not required. On the other hand, the Bayesian interpretation of range estimates may be preferable in all situations.

EARLY HISTORY

The first half of the 18th century saw religion versus science and mathematics argued in pamphlets by such theologians, philosophers and mathematicians as George Berkeley and David Hume. Hume, especially, argued that cause and effect can only be learned through observation, and not through tradition, beliefs or reasoning and logic. Although the essence of Hume's argument is still debated today, there is agreement that he believed that empirical evidence is required of correlated events for cause and effect to be established. Hume's treatise precipitated discomfort among the faithful who considered God to be the First Cause; you cannot observe a deity raising the sun each morning.

The Presbyterian Reverend Thomas Bayes attempted to counter Hume's claims mathematically. Amidst this environment of controversy, and using the fledgling math of probabilities, Bayes determined to find cause from observed effects rather than concluding the effects from an assumed cause; for example, perhaps it is sufficient to repeatedly observe the rising sun to conclude First Cause. Bayes began with a simple Boolean system to determine the position of an original ball on a table by observing whether a new ball thrown onto the table lands left or right of the original, while remaining blinded to the original ball's position. His system, which is not unlike Boolean search algorithms, had characteristics that we rely on today for causal analyses:

- A prior belief, with no corroborating evidence (data): the assumed initial position of the original ball
- Increasing amounts of evidence (data): each new ball's position relative to the original ball
- Updated belief (estimate): updated original ball's position given relative position of the new ball
- A resulting probability statement: probable position (region) of the original ball

This simple experiment reversed probabilities: rather than estimating the probability that a new ball would land left or right of the original ball, given the original ball's position, Bayes estimated the probability of the original ball's position using increasing amounts of observed data.

Whether through modesty or dissatisfaction with, or uncertainty about, his own work, Bayes did not promote or publish his findings. About a decade later, another Presbyterian Minister, Richard Price, published a refined version of Bayes' work on inverse probabilities in an attempt to prove the existence of a wise deity (the cause) by observing natural laws (the effects).

Independent of Bayes' and Price's work, and later in the 18th century, Pierre Simon Laplace, after reading Abraham de Moivre's *Doctrine of Chances* (the same book studied by Bayes), adopted the mathematics of probabilities to deal with variation in astronomical data relating to the positions of the planets and the sun. Laplace called his method the "probability of cause." In 1814, after decades of working with data, he published *Essai philosophique sur les probabilités* in which he specified principles of probabilities, one of which we would recognize as Bayesian probability: the probability of a cause given an event is proportional to the probability of the event given the cause.

Sharon Bertsch McGrayne writes a captivating history of Bayes' Theorem, from Bayes to Laplace to the present day. It is clear that Laplace laid the foundation for Bayes' Theorem and its application. The theorem, however, bears the name of his predecessor, a moniker that was first applied during the 1950s.

SUBJECTIVITY AND PRIORS

There are several ways to write conditional probabilities and Bayes' Rule, each being useful in a different context. Assuming A and B are observable events, $P()$ means probability, (A,B) means A and B occur together, and $(A|B)$ means A given that B has been observed, then:

$$P(A|B) = \frac{P(A,B)}{P(B)} \quad \text{or} \quad P(A,B) = P(A|B)P(B) = P(B|A)P(A) \quad (1)$$

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)} \quad \text{or} \quad P(A|B) = P(A) \frac{P(B|A)}{P(B)} \quad (2)$$

$$P(A|B) = \frac{P(B|A)P(A)}{\sum P(B|A_i)P(A_i)} \quad \text{or} \quad P(A|B) = P(A) \frac{P(B|A)}{\sum P(B|A_i)P(A_i)} \quad (3)$$

The algebraic inversion of probabilities in (1) gives rise to (2). Equation (3) stems from (2) by substituting the unconditional (marginal) probability of event B with the sum of its partitions over all possible outcomes of event A . Equations (2) and (3) are useful when we interpret Bayesian methods as updating prior probabilities with new evidence. An interpretation is:

My updated (posterior) belief of event A given that I have observed event B , $P(A|B)$, is dependent upon my prior belief of event A , $P(A)$, and the likelihood of event B given that event A has occurred, $P(B|A)/P(B)$.

For example, if event A represents a visit to a webpage with a SAS® advertisement, and B represents a purchase of SAS® software, then a reasonable question is whether the advertisement prompts viewers to purchase. The necessary pieces of information to estimate $P(\text{visit} | \text{purchase})$ include the probability that one visits the webpage, $P(\text{visit})$, and, for each purchase of software, whether the webpage was visited, $P(\text{purchase} | \text{visit})$ and $P(\text{purchase} | \text{no visit})$.

This example is one of cause and effect. In general, we are interested in $P(\text{cause} | \text{effect})$. In this example, the hypothesis of interest is that a visit to the webpage (and presumably seeing the advertisement), causes one to purchase the software; $P(\text{visit} | \text{purchase})$. Note that there is a temporal element in that the cause must precede the effect. The difficulty will be to assign the probability of visiting the webpage, $P(\text{visit})$, which is the prior probability to be updated with each purchase of software. How do you specify a prior, especially before any evidence or data are gathered; in this example, before anyone purchases the software?

When Laplace was made aware of Bayes' work (via Richard Price), he acknowledged Bayes' ingenuity in using an initial guess as to the position of the original ball on the table; i.e. before any observations were made. The subjectivity of the prior did not, presumably, cause Laplace any anxiety. Rather, he recognized that Bayes' initial estimate of the ball's position was needed to begin the learning algorithm of his theorem. Although the prior was subjective, the system learned with each iteration of observed data and corrected what may have been a poor estimate with a better estimate.

The subjective nature of priors has been a major complaint against Bayesian methods. However, frequentist methods are not devoid of subjectivity. During the planning of an experiment, effect sizes or variance estimates are needed to estimate sample size requirements. Inferences from hypothesis tests and confidence intervals rely on a pre-specified allowable type 1 error rate, which has been traditionally, and arbitrarily, set at 5%. In Bayesian analyses, you can use non-informative priors to mitigate the impact of subjectivity on the posterior estimates. Also, in operational systems in which priors are previous posteriors, subjectivity diminishes with time.

BELIEFS AND CREDIBILITY

There are a number of benefits to adopting Bayesian methods:

- Formal, mathematical method to use prior information
- Data are not required to specify a prior; can rely on subject matter expertise
- Ability to make probability statements about probable causes
- Ability to conduct scenario analyses (simulations) in complex cause and effect systems
- Ease of interpretation

Frequentist interpretation of probabilities is a ratio of long-run frequencies. That is, the number of times an occurrence is observed given the number of trials. Bayesian interpretation of probabilities is a degree of belief. When you read the weather forecast, do you interpret "30% chance of rain" as a frequency? That is, during previous similar weather patterns, it rained 3 out of 10 times? Or do you interpret the probability as the degree to which, on a scale of 0 to 100, forecasters believe it will rain on this given day? The degree of belief is a more natural interpretation in non-repeatable situations.

Things become more abstract in a statistical setting. Suppose you are estimating a regression parameter. If you know the distribution of the parameter, given the data, then you can make probability, or degree of belief, statements about the parameter. This is in contrast to a range estimate, or confidence interval, with

an associated statement about the “confidence” of long run frequencies. The difference between this two interpretations depends on what is considered to be random.

Modifying the table from this paper’s introduction:

$y = \alpha + \beta x$	Frequentist	Bayesian
Population parameter	β fixed but unknown	β random
Sample data	Random	Fixed
Inferences	$P\{Y, X \mid H_0(\beta=0)\}$ 95% CI: (β_L, β_U) random about fixed β	$P\{\beta \mid Y, X\}$ 95% Credible Region from distribution of β
Interpretation	Over repeated sampling, <u>an</u> estimated interval will capture β 95% of the time.	The probability that β is between β_L and β_U is 0.95.



Here is an example of a quadratic regression. Published estimates were used to establish the prior distributions for a Bayesian analysis with new experimental data. The frequentist column of estimates ignores the prior published estimates and relies solely on the new data.

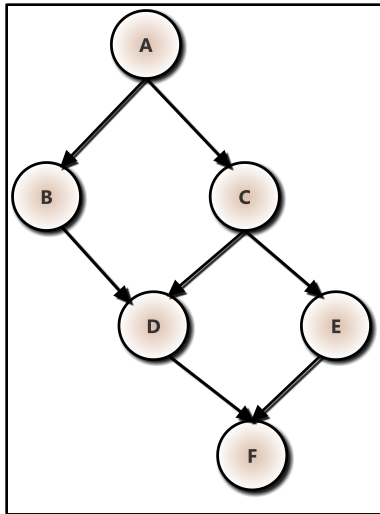
Parameter	Published estimates (priors)	Frequentist	Bayesian: Uniform Prior	Bayesian: Normal or Gamma Prior**
Intercept	24.60	22.54	22.51	23.24
Slope	6.06	7.56	7.57	7.33
Slope ²	-0.17	-0.28	-0.28	-0.27
Scale	10.90	12.91	13.10	12.50

** Normal used for coefficients, gamma used for the scale parameter

The 95% confidence interval for the slope parameter estimated using frequentist methods is 4.51 to 10.61. The interpretation of this interval is: “If I repeat this analysis with 100 independent samples, 95 of the estimated intervals for the slope parameter will include the true parameter value.”

The 95% Highest Posterior Density (HPD, see Bayesian Analyses and SAS below) for the slope parameter, using the informative priors, is 5.17 to 9.72. The interpretation for this interval is: “With 95% probability, the true value of the slope parameter is in this range.”

CAUSALITY AND BAYESIAN NETWORKS



A type of model that is gaining prominence is the Bayesian Network (BN). BNs are Directed, Acyclic Graphs (DAG) made up of nodes, representing random variables, and arcs (edges, links, or connectors) representing probabilistic dependencies. They help define, via visualization, the dependencies in a complex system or network and are so named because of three attributes:

- The node distributions and relationships may be subjective
- Bayes' Theorem is repeatedly used to update node distributions
- Inferences are of a causal nature, rather than correlations

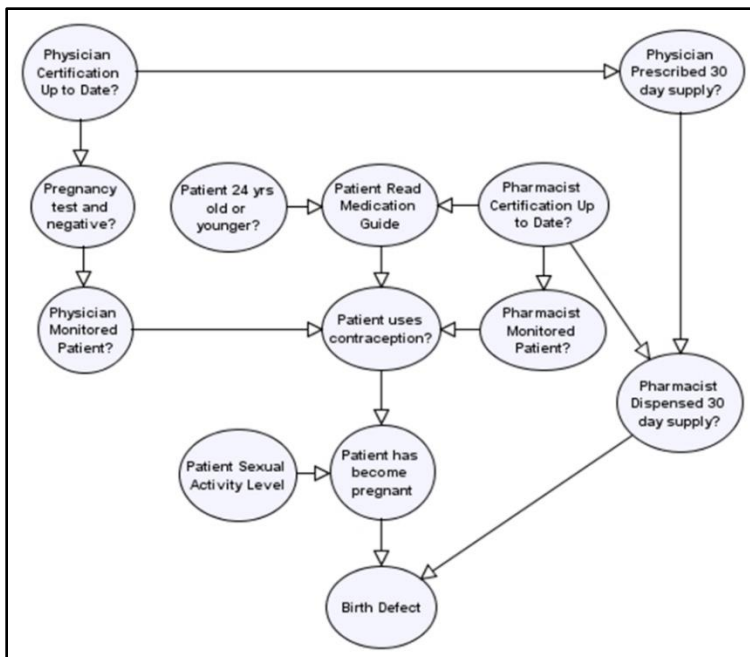
Bayesian probabilities are calculated for each node whenever new data become available, resulting in an updated probability of the occurrence of the outcome of interest, F , or a probable cause to F if it is known that F occurred. The interpretation of cause and effect derives from the single direction of the arcs (the "A" in DAG) and estimating the distributional dependencies of all random variables rather than just the target (as in

frequentist regression methods). In the DAG on the left, the arcs leaving A are diverging connectors, so A is a common cause to B and C . The arcs entering D are converging connectors, so D is a common effect of B and C .

Conditional independence and the Markov Property (Compatibility) of BNs greatly increase the efficiency with which distributions are updated. Conditional independence states that if $P(A,C) > 0$ and $P(A|B,C) = P(A|C)$, then events A and B are conditionally independent. That is, if it is possible that events A and C occur together, and knowing B does not provide any additional information to knowing C alone regarding the occurrence of A , then A and B are independent, conditional on knowing C . The Markov property states that, if a probability model correctly defines all of the conditional independencies represented by a BN, then that model and the BN are compatible. This is not unlike the assumption in regression that the specified model is correct.

Bayesian Networks have a natural application to risk management. They can incorporate subjective information from subject matter experts, assign probable cause in a scenario analysis or real occurrence of an undesirable event, and continually update the system as new information becomes available. Through

the assignment of probable cause, BNs indicate how limited resources should be allocated to mitigate risks of the undesirable event.



The U.S. Food and Drug Administration (FDA) has the option to require pharmaceutical firms to plan and implement a Risk Evaluation and Mitigation Strategy (REMS) as a requirement for market approval for a new drug. FDA's draft guidance to industry states that *Part of risk management activities should include an objective, evidence-based evaluation of the efficacy in risk communication*. The BN on the left represents a strategy to prevent birth defects to children of female patients prescribed a new drug. The strategy of the fictitious company hinges on certifying physicians and pharmacists to

prescribe and dispense the drug, their monitoring of patients, and a communication directly to the patient about the risks to unborn children. The effect of interest is a birth defect occurrence. The model includes possible causes as well as mitigation factors leading to the effect. The pharmaceutical firm can initiate the model using subject matter expertise (as did Bayes in 1763) and previously collected data. As the drug is prescribed and used in practice, the system updates factors' distributions to, in turn, update the probability of a birth defect. At any time, the pharmaceutical firm can analyze a scenario in which a birth defect occurred to assign probable cause. We can conduct a scenario analysis by setting the occurrence of a birth defect to "yes" and update all of the probability distributions. If, for example, the updated model indicates that the probability that the physician monitored the patient dropped by about 15%, which may have influenced the patient to relax her use of contraception, then to strengthen the risk mitigation program, more resources should be allocated to ensuring physicians monitor patients.

Financial institutions are also concerned with managing risks. The Basel Committee on Banking Supervision sets operational guidelines for financial institutions considered "too big to fail". Similar to FDA guidance, the Basel Committee publishes its recommendations and requirements. An example is *Principles for Effective Risk Data Aggregation and Risk Reporting*, in which Principle 7, in part, states:

Approximations are an integral part of risk reporting and risk management. Results from models, scenario analyses, and stress testing are examples of approximations that provide critical information for managing risk. While the expectations for approximations may be different than for other types of risk reporting, banks should follow the reporting principles in this document and establish expectations for the reliability of approximations (accuracy, timeliness, etc) to ensure that management can rely with confidence on the information to make critical decisions about risk. This includes principles regarding data used to drive these approximations.

Bayesian Networks are a tool that banks can use to remain compliant with Basel principles and to effectively allocate resources to mitigate operational risks.

BAYESIAN ANALYSES AND SAS

From *Introduction to Bayesian Analysis Procedures* in the SAS/STAT® 14.3 User's Guide.

SAS/STAT software provides Bayesian capabilities in six procedures: BCHOICE, FMM, GENMOD, LIFEREG, MCMC, and PHREG. The FMM, GENMOD, LIFEREG, and PHREG procedures provide Bayesian analysis in addition to the standard frequentist analyses they have always performed. ...The BCHOICE procedure provides Bayesian analysis for discrete choice models. The MCMC procedure is a general procedure that fits Bayesian models with arbitrary priors and likelihood functions.

All applications of Bayesian methods require that you specify a prior. A few attributes of priors include the following:

- All priors are subjective in that they represent a degree of belief.
- Priors that have little impact on the posterior (flat compared to the likelihood) are known as *objective, flat, diffuse, or noninformative*
- Priors are probability functions and therefore should integrate to one. If they do not, they are known as *improper* priors (e.g. uniform over the real line). However, they are sometimes useful.
- Priors may lead to *improper posteriors*, which cannot be used for inference
- Priors which are flat within the appreciable range of the likelihood (e.g. ± 2 or 3 sd of a normal mean) and have small values outside that range are said to be *locally uniform*. Jeffrey's prior is a useful prior in this category.
- *Conjugate* priors are priors that represent a family of distributions that lead to posteriors that are in the same family. That is, the posterior has the same distributional form as the prior. Examples include prior-likelihood combinations; Normal-normal, Beta-binomial, and Gamma-Poisson.

Inferences from Bayesian analyses rely on the posterior distribution of the parameter of interest. Point estimates can be the mean, median, or mode. Variances estimates arise from the posterior variance. Most

problems cannot estimate the posterior distribution using a closed-form solution. Rather, simulation algorithms are used. All Bayes applications in SAS rely on Markov chain Monte Carlo (MCMC) methods. You need to be careful to not confuse the estimated standard error for the estimated mean of interest and the MCMC Standard Error (MCSE) due to the simulation. The standard error is, in part, determined by the sample size and the MCSE is a function of the number of simulation iterations.

There are two types of Bayesian intervals; 1) equal-tailed intervals, $100(\alpha/2)^{\text{th}}$ to $100(1-\alpha/2)^{\text{th}}$, of the posterior estimate, and 2) Highest Posterior Density Regions (HDR or HPD), which is the shortest possible interval on the parameter in which the density of any point within the interval is higher than of any point outside the interval. HPDs are often preferred because they tend to be shorter intervals than the equal-tailed intervals.

Hypothesis testing is also possible with Bayesian methods, but is not typically the objective.

BAYESIAN NETWORKS IN ENTERPRISE MINER

Bayesian Networks are available in SAS Enterprise Miner 14.1 via the HP Bayesian Network (HPBN) node (and in Factory Miner 14.1). A synthetic data set, “Asia”, from Lauritzen and Spiegelhalter (1988) illustrates results from the HPBN node.

Shortness-of-breath (dyspnea) may be due to tuberculosis, lung cancer or bronchitis, or none of them, or more than one of them. A recent visit to Asia increases the chances of tuberculosis, while smoking is known to be a risk factor for both lung cancer and bronchitis. The results of a single chest X-ray do not discriminate between lung cancer and tuberculosis, as neither does the presence or absence of dyspnea.

The data set contains 5000 observations and 8 binary character variables whose values are either yes or no. Suppose that we want to determine the dependencies between a set of inputs, dyspnea, bronchitis, a visit to Asia, smoking, and a positive x-ray, and the verified presence of either tuberculosis or lung cancer.

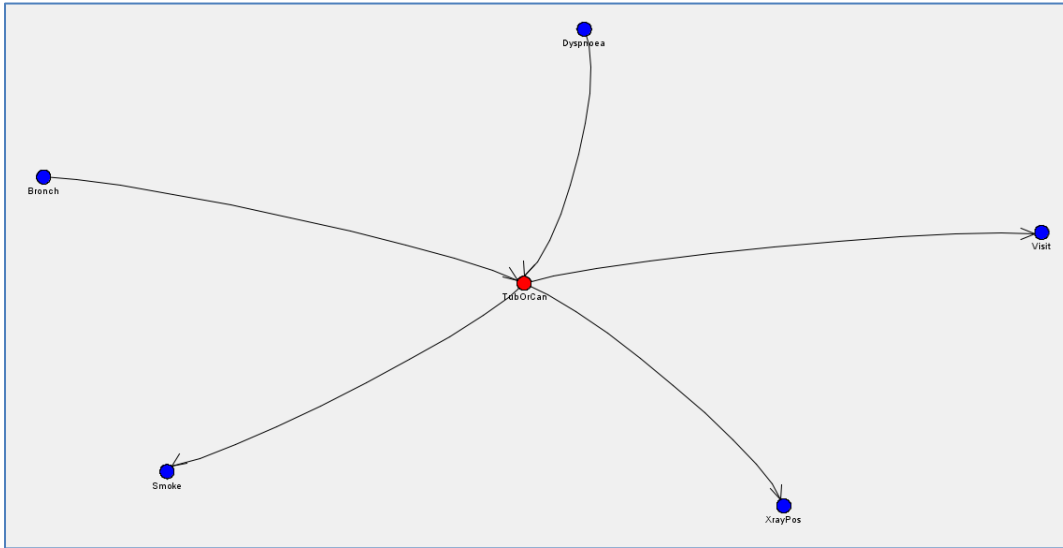
The Asia Data Set		
Variable	Label	Description
Dyspnea	Dyspnea	Presence of dyspnea?
Tuber	Tuberculosis	Presence of tuberculosis?
Cancer	Lung Cancer	Presence of lung cancer?
Bronch	Bronchitis	Presence of bronchitis?
Visit	Visit to Asia	Visit to Asia?
Smoke	Smoking	Smoker?
XrayPos	Chest X-ray	Positive chest X-ray?
TubOrCan	Tuberculosis or Lung Cancer	Presence of either tuberculosis or lung cancer?

We prepare and partition the data using other nodes preceding the HPBN node, including the metadata node to set the variable *TubOrCan* as the target. We configure the HPBN node to auto-select from four BN structure types and the possible models within each of those types. Possible models include not only whether a variable is in or out of the model, but the relationship between the nodes; be it a parent, child, or neighbor. HPBN fits the possible models using the Bayesian Information Criteria (BIC) and tests of independence and selects the best fitting model using the BIC on the validation data set.

For our implementation, we configure the HPBN node to not use the tuberculosis and cancer variables. We

also turn off variable pre-screening and selection because we want to explore the dependencies among the other inputs. Therefore, our possible models are reduced to the various BN structure types and the edges connecting the nodes. More details are provided in a companion SASGF paper, 2776-2018, by Wang and Amrhein.

The final model is a Parent-Child Bayesian Network. The target variable is always at the center and is shown as red. The inputs are blue and the direction of the arcs are from parent to child. Posterior probabilities are given as $P(\text{child} | \text{parent})$. You must consider the temporal relationship, which HPBN does not know, to declare probable cause for an effect.



Parent Node	Parent Condition	Child Node	Child Condition	Probability
TubOrCan	YES	Smoke	NO	0.154122
TubOrCan	YES	Smoke	YES	0.845878
TubOrCan	NO	Smoke	NO	0.524755
TubOrCan	NO	Smoke	YES	0.475245

For example, the *Smoke* node is a child of *TubOrCan*. The posterior conditional probability of smoking (cause) given tuberculosis or cancer (effect) is 0.846. Because we know that smoking pre-dates disease onset, **if this network completely**

represented all factors influencing disease onset, then we could conclude, with very high probability, that smoking causes tuberculosis or cancer.

Parent Node	Parent Condition	Child Node	Child Condition	Probability
Bronch	NO	TubOrCan	YES	0.018205
Dyspnoea	NO	TubOrCan	YES	0.018205
Bronch	NO	TubOrCan	NO	0.981795
Dyspnoea	NO	TubOrCan	NO	0.981795
Bronch	NO	TubOrCan	YES	0.314961
Dyspnoea	YES	TubOrCan	YES	0.314961
Bronch	NO	TubOrCan	NO	0.685039
Dyspnoea	YES	TubOrCan	NO	0.685039
Bronch	YES	TubOrCan	YES	0.063415
Dyspnoea	NO	TubOrCan	YES	0.063415
Bronch	YES	TubOrCan	NO	0.936585
Dyspnoea	NO	TubOrCan	NO	0.936585
Bronch	YES	TubOrCan	YES	0.097333
Dyspnoea	YES	TubOrCan	YES	0.097333
Bronch	YES	TubOrCan	NO	0.902667
Dyspnoea	YES	TubOrCan	NO	0.902667

The two parents of *TubOrCan*, bronchitis and dyspnea, may be symptoms of tuberculosis or cancer. Therefore, the posterior conditional probabilities could be used as diagnostic aids. For example, from the table at the left, a patient who presents with shortness of breath but without bronchitis has a 0.315 probability of having tuberculosis or cancer. Stated another way, tuberculosis or cancer causes shortness of breath with probability 0.315.

CONCLUSION

Bayesian methods have suffered a lack of use for centuries due to at least two issues; 1) the subjectivity inherent in the methods and the interpretation of probabilities as “degrees of belief” led many to characterize them as unscientific, and 2) applying the methods was limited to problems for which the posterior distribution had a closed form solution. Advances in computer processing, search algorithms, and simulation methods has addressed the second issue, leading to a resurrection of a debate of the first issue. The many advantages discussed in this paper have greatly increased the research and application of Bayesian methods.

This paper discussed the philosophy and thought process motivating Bayesian methods. The benefits to adopting Bayesian analytical methods includes the abilities to:

- Interpret inferences in a manner more closely aligned with natural decision making
- Incorporate prior information and knowledge
- Incorporate subject matter expertise without corroborating data
- Assign probable causes, rather than just correlations, to effects
- Simulate scenarios in complex cause and effect systems

Disadvantages include the need to:

- Assign distributions to the prior information
- Execute complex computations involving simulations
- Recognize situations in which the posterior distribution is heavily influenced by a subjective prior

REFERENCES & RECOMMENDED READING

1. Basel Committee on Banking Supervision. (2013). *Principles for effective risk data aggregation and risk reporting*. Retrieved from <http://www.bis.org/publ/bcbs239.pdf>
2. Lauritzen, S. L. and D. J. Spiegelhalter (1988). *Local Computations with Probabilities on Graphical Structures and Their Application to Expert Systems*. Journal of the Royal Statistical Society. Series B (Methodological), Vol. 50, No. 2, pp. 157-224. Blackwell Publishing for the Royal Statistical Society
3. McGrayne, S. B. (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*. New Haven: Yale University Press.
4. Pearl, J. (2009). *Causality: Models, reasoning, and inference* (2nd ed.). Cambridge: Cambridge University Press.
5. U.S. Department of Health and Human Services, Food and Drug Administration. (2009). *Format and content of proposed Risk Evaluation and Mitigation Strategies (REMS), REMS assessments, and proposed REMS modifications (Draft guidance for industry)*. Retrieved from <http://www.fda.gov/downloads/Drugs/.../Guidances/UCM184128.pdf>
6. Wang, Fei and John Amrhein (2018). *Bayesian Networks for Causal Analysis*. Proceedings of SAS Global Forum 2018, Denver Colorado, Paper 2776-2018. SAS Institute Inc., Cary, NC.

ADDITIONAL SUGGESTED READING

1. Bain, Robert. (2016). Are Our Brains Bayesian? *Significance*. 13:4, 14-19. The Royal Statistical Society. London.
2. Basel Committee on Banking Supervision. (2011). *Operational risk – Supervisory guidelines for the Advanced Measurement Approaches*. Retrieved from <http://www.bis.org/publ/bcbs196.pdf>
3. SAS Institute Inc. (2008). *SAS/STAT® 14.3 user's guide*. Cary, NC: SAS Institute Inc.
4. Savage, Leonard J. (1972). *The Foundations of Statistics*. (2nd revised ed.) New York: Dover Publications, Inc.

ACKNOWLEDGMENTS

Fei and I thank our fellow statisticians at McDougall Scientific Ltd. for their support and comments that improved this paper; Janet McDougall, Hong Chen, Jim Wang, and Blue Neustifter.

CONTACT INFORMATION

Your comments and questions are valued and encouraged. Contact the author at:

John Amrhein
McDougall Scientific Ltd.
jamrhein@mcdougallscientific.com

SAS and all other SAS Institute Inc. product or service names are registered trademarks or trademarks of SAS Institute Inc. in the USA and other countries. ® indicates USA registration.

Other brand and product names are trademarks of their respective companies.