Allen *Crime Science* 2014, **3**:2
http://www.crimesciencejournal.com/content/3/1/2

Crime Science
a SpringerOpen Journal

SHORT CONTRIBUTION Open Access

Investigative advising: a job for Bayes

Jared C Allen

Abstract

Background: Bayesian approaches to police decision support offer an improvement upon more commonly used statistical approaches. Common approaches to case decision support often involve using frequencies from cases similar to the case under consideration to come to an isolated *likelihood* that a given suspect either a) committed the crime or b) has a given characteristic or set of characteristics. The Bayesian approach, in contrast, offers formally contextualized estimates and utilizes the formal logic desired by investigators.

Findings: Bayes' theorem incorporates the isolated likelihood as one element of a three-part equation, the other parts being 1) what was known generally about the variables in the case prior to the case occurring (the scientific-theoretical priors) and 2) the relevant base rate information that contextualizes the evidence obtained (the event context). These elements are precisely the domain of decision support specialists (investigative advisers), and the Bayesian paradigm is uniquely apt for combining them into contextualized estimates for decision support.

Conclusions: By formally combining the relevant knowledge, context, and likelihood, Bayes' theorem can improve the logic, accuracy, and relevance of decision support statements.

Keywords: Investigative advising; Decision support; BIA; Bayesian statistics; Police investigations

Findings

Police investigators occasionally seek the support of specialists in various fields. Cases of murder and rape, for instance, prompt the need to utilize all available resources to prevent future offending by the perpetrators, and serial offenses (believed to have a single perpetrator) can prompt the employment of consultants to link crimes and anticipate likely sites of future offending (or the offender's "home base"; Rossmo 2000, 2009; Woodhams et al. 2007). The statistical training and specializations of academic criminologists and psychologists make them candidates for such consultancy (Alison and Rainbow 2011). In the United Kingdom (and some other Western countries) law enforcement agencies have such consultants on staff. The task of these professionals is referred to as Behavioral Investigative Advising (BIA).

The field of BIA is young and still establishing professional and scientific standards (Dowden et al. 2007; Alison and Rainbow 2011). The research literature and empirical basis of BIA are rapidly expanding and improving (Dowden et al. 2007; Almond et al. 2011). Investigators have reported that BIA consultancy is useful both as a second opinion and as a decision support tool (Rainbow 2011). This tool aims to be accurate, useful, specific, and falsifiable (Alison et al. 2003). This assures the consultancy is beneficial to police and allows for the product to be evaluated after the investigation.

The advising process can be summarized generally as using the knowns of an investigation to estimate unknowns useful to investigators; for example, moving from the known locations of a series of crimes to the possible residence or workplace of the offender (Rossmo 2000). BIA consultants can assist in locating, describing, and prioritizing suspects by contributing scientific knowledge and formal analysis of "national datasets and other relevant base rate data" (Rainbow et al. 2011 p. 37). That is, their contribution is the assimilation of research literature, evidence, and context to optimize decision making.

Due in part to its recent genesis as a scientific field of study, there are a multitude of quantitative approaches used by BIA professionals to arrive at estimates for decision support. The vast majority of these (e.g., correlation, Jaccard's indices, chi-square tests, logistic regression) may aptly be called "frequentist". That is, the majority of approaches involve either interpreting likelihoods from

Correspondence: jared.allen@psych.ryerson.ca
Department of Psychology, Ryerson University, Toronto, Canada

$$\text{(a) } P(\mathrm{H}\mid \mathrm{O}) = P(\mathrm{O}\mid \mathrm{H}) * P(\mathrm{H}) \,/\, P(\mathrm{O})$$

$$\text{(b) Posterior} = \text{Likelihood} * \text{Prior} \,/\, \text{Normalizing constant}$$

$$\text{(c) } P(\text{suspect is } x \mid \text{evidence}) = \frac{P(\text{evidence} \mid \text{suspect is } x) * P(\text{suspect is } x \text{ before analysis})}{P(\text{suspect is } x \text{ regardless of evidence})}$$

Figure 1 **Bayes' Theorem expressed in a) probability statements, b) Bayesian terms, and c) investigative language.**

frequency data or utilizing null hypothesis significance testing to interpret estimates of unknowns.

Bayesian statistical inference is the algorithmic combination of previous and new data to obtain the probability of one or more causes producing the new data (Gill 2009; de Morgan 1838). This is different from inferring the simple probability of said data being observed (randomly or otherwise), which is the cornerstone of the more commonly used frequentist methods.

Bayes' theorem formally combines quantifications of one's pre-analysis information (a *prior*), some base rate criminological and demographic data (a *normalizing constant*), and a *likelihood* of obtaining one's evidence. As shown in Figure 1, the prior and likelihood are multiplied together and divided by the normalizing constant, yielding one's new conclusion or estimate (the *posterior*). This is more generally expressed as: The probability of a hypothesis (H) given an observation (O) is equal to the probability of obtaining the observation given the hypothesis is true, multiplied by the prior probability of the hypothesis, divided by the unconditional probability of obtaining the observation.

Key distinctions between Bayesian and frequentist (also called Fisherian) approaches to BIA estimation are the use of a null hypothesis and the use of prior information. Bayesian logic involves treating data as constant and modelling one's belief about relationships in the data based on the context of the data and the data, whereas frequentist logic involves treating the data as random, ignoring the context of the information so as to be

Table 1 Differences between Bayesian and Frequentist/Fisherian approaches to investigative inference

	Bayesian	Frequentist/Fisherian
Context	Incorporates past knowledge	Ignores past knowledge
Null hypothesis	Result based on strength of the evidence	Result typically (but not necessarily) based on assumption of no effect or assumption of a statement counterfactual to one's question
What is random	The parameters describing the relationships within the data are treated as random within some distribution. (e.g., in Markov chain Monte Carlo methods, the data is treated as constant, but the relationships taking the researcher from the data to a prediction are randomly iterated to optimize the model for each data value and determine how parameter values vary)	The data are treated as random so that the likelihood of obtaining it under the null can be assessed
Logic	Follows "inverse logic", moving from effect to estimation of cause	Typically uses null logic: rejection of no effect to infer effect
Philosophy	Probability is a measure of evidence, belief, or willingness to gamble based on all available information	Probability is relative frequency over time.
Summative statement	"The probability of H, given the evidence, is x%"	"If its contrary were true, then the chances of H (or a more extreme statement of H) would be less than x%"
Primary difficulty	New information must compete with old, making the process of discovery more conservative and necessarily cumulative	The assumption of no difference is always false. Given a large enough sample size, any difference will be found statistically significant.
Pragmatic difficulty for BIA	Determining the measure of one's priors can be difficult, and Bayesian methods can be perceived as unscientific, especially in legal circles	Does not produce estimates of the form typically desired (e.g., "a 77% chance"), and results logically pertain to the data itself, not to the prediction of new cases

"objective", and—typically—evaluating the existence of a relationship from the initial standpoint of the assumption that no relationship exists. Table 1 details key relevant differences between Bayesian and frequentist approaches to statistical inference. Note, however, that some exceptions to these differences may exist, especially when considering very simple applications of Bayes' theorem and very complex applications of frequentist statistics.

Bayes' theorem can be effective both as a tool and as an analogue to the logical problems faced by investigators. Tartoni et al. (2006) note that Bayesian analysis is well-suited for nearly all aspects of forensic investigation, and Schneps and Colmez (2013) illustrate the grievous errors that can occur when cases are built solely based on an isolated frequentist analysis of the evidence. For example, calculating a simple 1 in 6 chance of identifying an offender from a line-up versus a 1 in 12 chance may lead one to believe that having more individuals as foils in a police line-up increases the posterior probability that an accurate match was made. Wells and Turtle (1986) noted that this is not the case. They also shed empirical light, using a Bayesian updating model, on the practice of having all-suspect line-ups, which they found increases the risk of false identification.

Blair and Rossmo (2010) tackle the issue of assigning prior probability values for decision support. They argue

Table 2 Procedural comparisons based on a (highly simplified) investigative advising example

Example case

Given: Two homicide cases in which knives and strangle wires were used (i.e., a knife and strangle wire were used in case 1 and a knife and strangle wire were used in case 2).

Task: Assess whether

a) the two cases are linked (i.e., they have a common offender), and

b) the offender was known or a stranger to the victims.

	a) Case linkage	**b) Offender characteristic**
Dimensional frequentist approach	1) "Crunch" all data from a relevant database into a minimal number of fundamental dimensions	1) The dimensional scores of the cases (obtained for "a") point vaguely to certain offender characteristics that belong to or have similar dimensional scores as the cases themselves (e.g., given the offender used both a knife and a strangle wire, this may yield a higher score on a "sadism" dimension. Assume being a stranger offender is associated with sadism: If the offender is a stranger, then the evidence is more likely than the evidence would be if the offender were not a stranger).
	2) Link the cases based on the similarity of their scores along these dimensions such that, if the cases have uncommonly similar dimensional scores based on the frequencies of such scores (according to some predetermined rule), it is predicted that they are linked.	2) Use more specific base rate analysis to obtain pared-down (quantified) likelihood estimates of the offender being a stranger by seeing what percentage of homicide cases involving a knife and strangle wire also involved a stranger offender (this number, the pared-down base rate, would constitute the likelihood estimate).
	Note that this analysis estimates how probable the *scores* are assuming they occur by chance only, which is a different question than whether they are indeed linked.	1) Narratively combine the above to obtain 1) an argument, and 2) a quantification.
Bayesian approach	1) Keep each behavioural variable (in both the database and the cases themselves) as an individual unit of information, and evaluate the case information using Bayesian reasoning. For this, iteratively train a model with the cases of a relevant database to predict the random variable: linkage.	1) Obtain the prior likelihood of the offender being a stranger to the victim (this could be the simple base percentage of stranger homicides among all homicides, or an investigator's initial opinion).
	2) Produce a probability estimate of whether the cases are linked given their behavioural variable values. That is, combine using Bayes' theorem the case information and the trained model developed from the database, into a posterior estimate. This approach treats the conditional likelihood (from "a 2" above) as only one element of the linkage estimate.	2) Produce a conditional likelihood, based on the database, of an offender using a knife and wire given the offender is a stranger to the victim.
		1) Combine the prior, likelihood, and the case data using Bayes' theorem. In this way, the probability that the offender is a stranger to the victim, based on the fact that the offender used a knife and wire, can be explicitly assessed within the context of the (specific) pertinent data, and a singular value can be obtained.

that a Bayesian approach can improve estimation of guilt, and suggest assigning probability ranges to single or multiple pieces of evidence. They note that this does not solve the problem of assigning "guilt" values to pieces of evidence, but the approach can result in "more systematic assessments and improved investigative decision making" (Blair and Rossmo 2010 p. 133). On a cautionary note, when using databases of convicted criminals to estimate guilt, both the Bayesian and frequentist statistical approaches may perpetuate biases in a system of justice. That is, using the "usual suspects" to predict characteristics of offenders could lead to further focus on these individuals at the expense of other potential investigative leads. The Bayesian approach is not immune to this criticism, though it is less vulnerable to the specific claim that its inherent logic is biased to this conclusion. Frequentist approaches assume the validity of a null hypothesis, that is, they assume the predictor and outcome variables may legitimately be thought to not be related. When this logic is used to evaluate a candidate suspect whose prior offenses are used in the model quantifying his guilt, this assumption is grossly violated and the logic of the frequentist estimator is circular. That is, the offender's statistical relationship to himself is used as evidence against him because the test, in assuming no relationship, finds his relationship to himself "significant". In frequentist approaches, this is a violation of the logic of the method. In Bayesian approaches this is not a logical violation (since no null assumption is required and the context of the information is adequately incorporated). However, the potential for an offender's resemblance to himself to make his candidacy as a suspect more likely still remains. The potential for this concern should be considered when using *any* statistical method to parse local databases for BIA consultancy.

Table 2 presents a procedural comparison of two approaches to investigative advising, taken from Salo et al. (2012) and Allen et al. (in press). These papers empirically compare Bayesian to non-Bayesian prediction for investigative advising. Salo et al. (2012) informs column a. The study compared use of a Bayesian updating model with a dimensional model to link homicide cases using only offender behavioural information (i.e., only details of what the offender did). Both models utilized identical real-world data. The Bayesian approach, by better accounting for absent information, resulted in 83.6% of cases being correctly classified, versus 62.9% by the dimensional approach. Allen et al. (in press) informs column b. The study compared an empirical Bayesian approach to a "pared-down" base rate method of estimating offender characteristics. The Bayesian approach, by incorporating more contextual information, resulted in 74.6% prediction accuracy versus 63.5% accuracy of the base rate method.

Bayesian methods are subject to a disproportionate amount of criticism for being "subjective" and prone to misuse (e.g., Doren 2006). This is due in part to the forthright philosophy of Bayesian analysis, which formally "confesses" that Bayesian estimates, *like all other estimates*, are a product of, and representative of, beliefs about the hypothesis being explored. Popperian objectivity requires that the statements and evidence be entirely in observable space (Popper 1972). Therefore, provided all the values used in an analysis are thoroughly explained and justified, Bayesian methods are no less objective than their frequentist counterparts (which involve many subjective choices).

Bayesian methods can formally contextualize, and thus improve, frequentist analysis. In the 20th century, insurance companies used Bayesian inverse probability, contrary to a rabidly Fisherian zeitgeist, without knowing that their computations were incorporating Bayes' theorem (McGrayne 2011). Similarly, courts in the United States have been using Bayesian risk assessments (Donaldson and Wollert 2008; Wollert 2007) while also lambasting Bayesian approaches (e.g., Doren 2006). Conversely, BIA research has largely used frequentist methods to perform a fundamentally Bayesian task. Whatever the reputation of Bayesian analysis, the task and field of BIA are fundamentally Bayesian. A Bayesian approach to investigative advising is therefore the most logical and promising way forward.

Abbreviation

BIA: Behavioural investigative advising.

Competing interests

The author declares he has no competing interests.

Acknowledgements

Thank you to all five reviewers and the editorial staff, with special credit to Reviewer 4 for improving the manuscript's technical rigor. This research was funded in part by the Social Sciences and Humanities Research Council of Canada.

Received: 2 August 2013 Accepted: 11 November 2013
Published: 7 April 2014

References

Alison L, Rainbow L (ed) (2011) Professionalizing offender profiling: forensic and investigative psychology in practice. Routledge, London

Alison L, Smith MD, Eastman O, Rainbow L (2003) Toulmins philosophy of argument and its relevance to offender profiling. Psychol Crime Law 9(2):173–183

Allen JC, Goodwill AM, Watters K, Beauregard E (in press) Base rates and Bayes' theorem for decision support. Policing: An Int J Police Strateg Manage

Almond L, Alison L, Porter L (2011) An evaluation and comparison of claims made in behavioural investigative advice reports compiled by the National Policing Improvement Agency in the United Kingdom. In: Alison L, Rainbow L (ed) Professionalizing offender profiling: forensic and investigative psychology in practice. Routledge, London, pp 250–263

Blair JP, Rossmo DK (2010) Evidence in context: Bayes' theorem and investigations. Police Q 13:123–135

De Morgan A (1838) An essay on probabilities and their application to life contingencies and insurance offices. Longman, Orme, Brown, Green, & Longmans, London

Donaldson T, Wollert R (2008) A mathematical proof and example that Bayes's theorem is fundamental to actuarial estimates of sexual recidivism risk. Sex Abuse 20(2):206–217

Doren DM (2006) Battling with Bayes: when statistical analyses just won't do. Sex Offender Law Report 7(4):49–50. 60–61

Dowden C, Bennell C, Bloomfield S (2007) Advances in offender profiling: a systematic review of the profiling literature published over the past three decades. Journal of Police and Criminal Psychology 22:44–56

Gill J (2009) Bayesian methods, a social and behavioural sciences approach, 2nd edition. CRC Press, London

McGrayne SB (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, New York

Popper K (1972) Objective knowledge: an evolutionary approach. Oxford University Press, London

Rainbow L (2011) The UK approach to the management of behavioural investigative advice. In: Alison L, Rainbow L (ed) Professionalizing offender profiling: forensic and investigative psychology in practice. Routledge, London, pp 5–17

Rainbow L, Almond L, Alison L (2011) BIA support to investigative decision making. In: Alison L, Rainbow L (ed) Professionalizing offender profiling: forensic and investigative psychology in practice. Routledge, London, pp 35–50

Rossmo DK (2000) Geographic profiling. CRC Press, New York

Rossmo DK (2009) Geographic profiling in serial rape investigations. In: Hazelwood RR, Burgess AW (ed) Practical aspects of rape investigation: a multidisciplinary approach, 4th edition. CRC Press, Boca Raton, pp 139–170

Salo B, Sirén J, Corander J, Zappalà A, Bosco D, Mokros A, Santtila P (2012) Using Bayes' theorem in behavioural crime linking of serial homicide. Leg Criminol Psychol. Advance online publication. doi:10.1111/j.2044-8333.2011.02043.x

Schneps L, Colmez C (2013) Math on trial: how numbers get used and abused in the courtroom. Basic Books, New York

Tartoni F, Aitken C, Garbolino P, Biedermann A (2006) Bayesian networks and probabilistic inference in forensic science. John Wiley & Sons, Ltd, New York

Wells GL, Turtle JW (1986) Eyewitness identification: the importance of lineup models. Psychol Bull 99(3):320–329

Wollert R (2007) Poor diagnostic reliability, the null-Bayes logic model, and their implications for sexually violent predator evaluations. Psychology, Public Policy, and Law 13(3):167–203

Woodhams J, Bull R, Hollin C (2007) Case linkage-identifying crimes committed by the same offender. In Kocsis (Ed.), Criminal profiling: International theory, research, and practice (pp. 117–133). Humana Press Inc., Totowa, NJ

doi:10.1186/2193-7680-3-2
Cite this article as: Allen: **Investigative advice: a job for Bayes.** *Crime Science* 2014 **3**:2.