THE REASONING RULES OF (FORENSIC) SCIENCE

Paolo Garbolino[1]

Department of Architecture and Arts, IUAV University of Venice, Italy

***Abstract**. Probability theory can be interpreted as a system of rules for coherent behaviour in un uncertain environment. Bayes' theorem is a rule for making educated guesses and learning from experience, and it is a fundamental tool of scientific reasoning. The odds form of the theorem offers a way of measuring the relative weight of support given by evidence to alternative hypotheses that fits very well with the main task of forensic science.*

***Keywords**: Coherence, Bayes' theorem, Odds ratio, Weight of evidence.*

1. PROBABILITY AS A LOGICAL CALCULUS

Forensic reasoning is a kind of scientific reasoning and it must follow the 'good rules' of scientific method. The 'scientific method' is a qualified and systematic extension of rational behaviour. In everyday life we have to cope with a complex environment, and we are continuously required to make educated guesses, and developing uncertain inferences, in order to act in a coherent way in that environment. The Bayesian paradigm claims that the fundamental normative rule of uncertain reasoning in everyday life, and in scientific practice, is Bayes' theorem.

It is well known that the addition law of probability theory can be proved a standard of coherence for the degrees of belief held by a person, by means of the so-called *Dutch Book theorem* (Ramsey, 1931; de Finetti, 1937). From a subjective Bayesian point of view, also the product rule is not a definition but a standard of coherent reasoning under uncertainty, which can be as well justified by a *synchronic Dutch Book* argument, in terms of conditional bets (de Finetti, 1937; Jeffrey, 1988 and 2004). Both are rules for combining sets of degrees of belief held by a person at the same time, upon a given body of evidence.

The product rule immediately provides the way for updating a person's degrees of belief, when the body of evidence changes upon time, if the so-called *principle of conditionalization* is accepted. The principle says that your new subjective probability distribution Q, based on new evidence E, ought to be equal to your old conditional probability distribution P given E, provided that $P(E) > 0$:

[1] Paolo Garbolino Corresponding Author, email: pgarboli@iuav.it

$$Q(-) = P(-\,|E). \tag{1}$$

Given that a *Conditional Dutch Book theorem* can be proved also for the principle of conditionalization (Teller, 1973; Jeffrey, 1988; Skyrms, 1990), Bayesians accept the principle as a rule of coherence reasoning under uncertainty.

The philosopher Richard Jeffrey (Jeffrey, 1983) has given a generalization of the principle. Suppose that the new information E is not known with certainty, that is, $Q(E) \neq P(E)$ and $Q(E) < 1$; then we can rewrite (1) as follows:

$$\frac{Q(-\,\&\,E)}{Q(E)} = \frac{P(-\,\&\,E)}{P(E)} \tag{2}$$

and thus

$$Q(-\,\&\,E) = \frac{Q(E)}{P(E)} P(-\,\&\,E). \tag{3}$$

Given that you can update $Q(-\,\&\,\bar{E})$ in the same way, your new probability distribution will be:

$$Q(-) = Q(-\,\&E) + Q(-\&\,\bar{E}). \tag{4}$$

This generalization is the logical core of the updating algorithms implemented in the inference engines of *Bayesian networks* (Taroni et al., 2006).

Bayes' theorem is a straightforward consequence of the product rule and the additivity rule and it provides the machinery that can put at work the principle of conditionalization. It can be useful to write the theorem in the following way, because it immediately captures the logical relationships between a given hypothesis H and a body of evidence E (Polya, 1954):

$$P(H|E)P(E) = P(E|H)P(H). \tag{5}$$

It can be easily seen that:

(i) the probability of the hypothesis H, upon knowledge of information E, usually called its *posterior* probability, is greater than the probability of the same hypothesis *without* knowledge of E, usually called its *prior* probability, if and only if the *likelihood* of H given E, i.e. $P(E|H)$, is greater than the absolute probability of E;

(ii) in such a case, given that by convexity of probability measures it holds:

$$P(E|\bar{H}) \leq P(E) < P(E|H) \tag{6}$$

we see that evidence E supports hypothesis H if E is *unexpected*, or *surprising*, *unless H is true*; and that, the less probable evidence E is, unless H is true, the more informative is *E* about *H*. Therefore, Bayesians agree with Karl Popper's recommendation that good tests of a scientific theory must look for improbable predictions, unless the theory is true, and that the more improbable the prediction is, the better the test (Popper, 1959);

(iii) Popper's falsification rule is a particular case of Bayes' theorem, that applies when *H* implies *E*, because in that case $P(\bar{E}|H) = 0$, and $P(H|\bar{E}) = 0$;

(iv) genuine empirical hypotheses *H* can never be verified, because $P(H|E) = 1$ if and only if $P(H) = P(E)$, but this cannot be the case if *E* is empirical evidence, because there will be always a possible alternative explanation of an empirical fact. Therefore, we must always allow for $P(E|\bar{H}) > 0$, even though we might not be able to fully work out the alternative hypothesis and calculate the relative likelihood. This 'always possible' alternative is what philosophers of science sometimes call the 'catch-all hypothesis'.

Fortunately, we need neither to calculate the likelihood of an unknown hypothesis nor to estimate its prior probability, in order to be able to use Bayes' theorem as a rule of inference.

2. THE BALANCE OF PROBABILITIES

Psychologists tell us that one of the more entrenched biases of common sense reasoning is to focus on one single hypothesis, without taking into account alternatives (Kahneman, 2011). The main lesson philosophers and historians of science have drawn by Thomas Kuhn's *The Structure of Scientific Revolution* (Kuhn, 1970), is that no scientific theory has been rejected simply because there is some evidence which 'falsifies' it, but it is rejected only if an alternative is available for explaining away the 'anomalies'. Bayesian epistemology is able to provide a general account of scientific practice (Bovens and Hartmann, 2003; Howson and Urbach, 1996), and to teach common sense reasoning that a correct assessment of evidence always requires a comparison of at least two competing hypotheses.

Suppose we consider testing two *mutually exclusive*, but not *exhaustive*, hypotheses as, for example, two simple statistical hypotheses ($\theta = \theta_0$; $\theta = \theta_1$): the comparison makes use of the *odds form* of Bayes' theorem. The *prior odds* on θ_0 *against* θ_1, based on the evidence *x* we have before the test, are:

$$\frac{P(\theta_0 \mid x)}{P(\theta_1 \mid x)}. \tag{7}$$

The *posterior odds* on θ_0 *against* θ_1, based on the observation of the result *y* of the test, are given by the formula:

$$\frac{P(\theta_0 \mid x,y)}{P(\theta_1 \mid x,y)} = \frac{P(y \mid \theta_0,x)}{P(y \mid \theta_1,x)} \times \frac{P(\theta_0 \mid x)}{P(\theta_1 \mid x)}. \tag{8}$$

The ratio $P(y|\theta_0, x)/P(y|\theta_1, x)$ is the *likelihood ratio* and it is called the *Bayes factor*: it is the measure of the *relative strength of support* which evidence *y* gives to θ_0 *against* θ_1, given evidence *x*. Given the assumption, usually made for probabilistic models, that all the observations are independent, conditional on each particular value of the parameter θ, formula (8) simplifies in:

$$\frac{P(\theta_0 \mid x,y)}{P(\theta_1 \mid x,y)} = \frac{P(y \mid \theta_0)}{P(y \mid \theta_1)} \times \frac{P(\theta_0 \mid x)}{P(\theta_1 \mid x)}. \tag{9}$$

No 'catch-all hypothesis' appears in (9), and we have all the data to make the necessary calculations.

The price we have to pay is that no quantitative posterior probabilities can be calculated for θ_0 and θ_1, when they are not exhaustive hypotheses.

> "In fact, most applications of Bayesian standpoint in everyday life, in scientific guessing, and often also in statistics, do not require any mathematical tool nor numerical evaluations of probabilities; a qualitative adjustment of beliefs to changes in the relevant information is all that may be meaningfully performed" (de Finetti, 1974, p. 117).

Suppose that, prior to the observation of evidence y, you believe that hypothesis θ_0 is more credible than hypothesis θ_1. Then, it is easy to see from (9) that the posterior probability of θ_1 will be greater than the posterior probability of θ_0, that is, your preferences should be changed from θ_0 to θ_1, if and only if the likelihood ratio of θ_1 to θ_0 is greater than the prior odds of θ_0 against θ_1 (Salmon, 1990):

$$\frac{P(y \mid \theta_1)}{P(y \mid \theta_0)} > \frac{P(\theta_0 \mid x)}{P(\theta_1 \mid x)}. \tag{10}$$

In case where we can have a list of exclusive *and* exhaustive hypotheses, a decision-theoretic approach can easily make use of the odds form of Bayes' theorem. Suppose, for sake of simplicity, that there are only two exhaustive hypotheses θ_0 to θ_1, and that a *loss function L* can be estimated, so that a 'loss' is incurred when the false hypothesis is chosen, and there is no 'loss' when the true hypothesis is chosen.

Table 1: A decision matrix

action	θ_0	θ_1
a_0: choosing θ_0	0	L_{01}
a_1: choosing θ_1	L_{10}	0

The *expected loss EL* of choosing hypothesis θ_i on data *x*, y, is:

$$EL(a_i|x,y) = L_{ij}(\theta_j|x,y) \tag{11}$$

and the rational decision is to take the action with the *lowest expected loss*, decision that involves a comparison of posterior odds with the 'losses' possibly incurred in choosing the wrong hypothesis. Then, hypothesis θ_1 will be preferred to hypothesis θ_0 if and only if:

$$\frac{P(\theta_1 | x,y)}{P(\theta_0 | x,y)} > \frac{L_{10}}{L_{01}} \tag{12}$$

that is:

$$\frac{P(y|\theta_1,x)}{P(y|\theta_0,x)} \times \frac{P(\theta_1|x)}{P(\theta_0|x)} > \frac{L_{10}}{L_{01}}. \tag{13}$$

Rewriting (13), we obtain that θ_1 will be preferred to θ_0 if and only if:

$$\frac{P(y|\theta_1,x)}{P(y|\theta_0,x)} > \frac{L_{10}P(\theta_0|x)}{L_{01}P(\theta_1|x)}. \tag{14}$$

The loss ratio in (13) fixes a threshold for posterior odds, also called the *posterior odds cutoff*, and we can see that (10) is the particular case where the loss is symmetric: $L_{10} = L_{01}$.

3. THE BALANCE OF JUSTICE

According to Bayesian epistemology, probability always refers to a single event, and there is only one way of reasoning for scientific theories in general, statistical hypotheses, and *forensic hypotheses*. The task of forensic scientists is to estimate the probabilities of the occurrence of unique events and to assess evidence E on the light of two alternative hypotheses, the prosecution's hypothesis *Hp* and the hypothesis proposed by the defence *Hd*. Therefore, what the scientist is asked to evaluate is measured by the likelihood ratio of the prosecution hypothesis:

$$\frac{P(E \mid Hp)}{P(E \mid Hd)}. \tag{15}$$

A useful function of the likelihood ratio is its logarithm that has been called by Good the *weight of evidence* (Good 1950; 1988) because in such a way the relative support provided by evidence enjoys the additive property required by an appropriate information function. Posterior odds

$$\frac{P(Hp \mid E)}{P(Hd \mid E)} = \frac{P(E \mid Hp)}{P(E \mid Hd)} \times \frac{P(Hp)}{P(Hd)} \tag{16}$$

can be rewritten as

$$log\left[\frac{P(Hp \mid E)}{P(Hd \mid E)}\right] = log\left[\frac{P(E \mid Hp)}{P(E \mid Hd)}\right] + log\left[\frac{P(Hp)}{P(Hd)}\right]. \tag{17}$$

The weight of evidence associated with the likelihood ratio is the additive change to the prior odds, due to evidence E only. Odds vary from 0 to ∞, while their logarithms vary from $-\infty$ to $+\infty$.

The idea of using the log odds as a measure of the 'weight of evidence' is due to Peirce (Peirce, 1878), but his definition applies to the special case where the prior odds are 1, i.e. $P(H) = P(\bar{H}) = 0.5$; in such a case, the prior log odds are 0 and the weight of evidence is equal to the posterior log odds. A similar notion of the '*weight of argument*' was put forward by Keynes to indicate the absolute amount of *relevant* evidence, independently whether or not this evidence is positively, or negatively relevant, i.e., whether or not it raises the probability of a hypothesis. The mathematical properties this measure was supposed to satisfy (Keynes, 1921, pp. 78-79, 84) mixing together two different modern concepts of *information* (Hilpinen, 1970), namely, the concept of *semantical information* (Popper, 1959) which refers to the logical content of a proposition, and the concept of *entropy* used in communication theory (Shannon and Weaver, 1949), which refers to the expected information.

The concept of entropy has been used to measure the amount of *expected information* before an experiment (Lindley, 1956), allowing the calculation of the *expected value of information*. Consider again Table 1 above: the best action a_i is that one for which the expected loss

$$\sum_{j=1}^{2} L_{ij} P(\theta_j) \tag{18}$$

is minimized. Now, if you know the true hypothesis, the action is that one with the smallest loss in the column corresponding to the true hypothesis. Therefore, to calculate the *expected loss with perfect information* you must multiply the minimum loss for each hypothesis by the probability of that hypothesis, and sum all these products:

$$\sum_{j=1}^{2}\left(min_i L_{ij}\right)P\left(\theta_j\right). \tag{19}$$

Before knowing which hypothesis is true, you would have chosen an action according to formula (18): therefore, the difference between (18) and (19) measures the reduction of the expected loss or, equivalently, the *expected value of perfect information*. Notice that this measure satisfies one of the *desiderata* of Keynes' 'weight', namely, that an increase in the relevant evidence available is positive, independently from the fact that probability is raised or lowered. Indeed, the expected value of perfect information is always greater than zero, because, whatever action is taken without perfect information, the value of (18) will be greater than the value of (19), since every loss L_{ij} in the former is replaced by a loss ($min_i L_{ij}$) in the latter, which cannot be greater.

The problem of the 'catch-all hypothesis' does exist, in principle, also for the evaluation of forensic evidence, for there will be always many hypotheses available to explain the occurrence of empirical facts as the production of physical traces as, for example, blood stain on the ground. It is true that a proposition like 'the suspect did not commit the fact' is the logical negation of the basic prosecution's proposition, but in order to be able to evaluate likelihood ratios more specific propositions are needed. In legal settings the probability of the 'catch-all hypothesis' can be considered so low that the prosecution and the defense hypotheses can be safely taken *as if* they were exhaustive, and on that assumption, and only on that assumption, we are allowed to pass from likelihood ratios and posterior odds to posterior probabilities. In practice, a comparative evaluation of odds is what can be reasonably done.

> "The aspiration of the legal system is to approach an assessment of odds. The means by which this is done in the vast majority of cases is to consider the two parties' respective hypotheses. It can be shown that a good approximation of the probability of a hypothesis can usually be attained by comparing it with the next most likely hypothesis. On the assumption that the two most likely hypotheses in a legal case are those advanced by the respective parties, this is what the legal system does" (Robertson and Vignaux, 1993, pp. 471-472).

If we apply the odds form of Bayes' theorem to the two parties' hypotheses

$$\frac{P(Hd \mid E)}{P(Hp \mid E)} = \frac{P(E \mid Hd)}{P(E \mid Hp)} \times \frac{P(Hd)}{P(Hp)} \tag{20}$$

we can realize that our legal system requires that, in tort law, the prior odds must be 1: thus, in order to give a verdict for the plaintiff, it is enough that $P(E|Hp) > P(E|Hd)$ (the *preponderance of evidence* standard of proof). But in criminal law, our legal system requires that the prior odds of *Hd* against *Hp* are extremely high and that posterior odds must be completely outbalanced to give a verdict for the prosecution (the *beyond any reasonable doubt* standard of proof): thus, the likelihood ratio $P(\mathrm{E}|Hp)/P(E|Hd)$ must be, accordingly, extremely high.

REFERENCES

Bovens, L. and Hartmann, S. (2003). *Bayesian Epistemology*. Clarendon Press, Oxford.

De Finetti, B. (1937). La prévision: ses lois logiques, ses sources subjectives. In *Annales de l'Institut Henri Poincaré*. 7: 1–68.

De Finetti, B. (1974). Bayesianism: its unifying role for both the foundations and applications of statistics. In *International Statistical Review*. 42: 117–130.

Good, I.J. (1950). *Probability and the Weighting of the Evidence*. Griffin, London.

Good, I.J. (1988). The interface between statistics and philosophy of science. In *Statistical Science*. 4: 386-397.

Hilpinen, R. (1970). On the information provided by observation. In J. Hintikka and P. Suppes, editors, *Information and Inference*, 97-122. Reidel, Dordrecht.

Howson, C. and Urbach, P. (1996). *Scientific Reasoning: The Bayesian Approach*. 2nd ed. Open Court, Chicago.

Jeffrey, R.C. (1983). *The Logic of Decision*. 2nd ed. University of Chicago Press, Chicago.

Jeffrey, R.C. (1988). Conditioning, kinematics and exchangeability. In B. Skyrms and W.L. Harper, editors, *Causation, Chance and Credence, I*, 221-255. Kluwer, Amsterdam.

Jeffrey, R.C. (2004). *Subjective Probability: The Real Thing*. Cambridge University Press, Cambridge.

Kahneman, D. (2011). *Thinking, Fast and Slow*. Farrar, Straus and Giroux, New York.

Keynes, J.M. (1921). A treatise on probability. In E. Johnson and D. Moggridge, editors, *The Collected Writings of J. M. Keynes*, volume *VIII*. Cambridge University Press, London.

Kuhn, T.S. (1970). *The Structure of Scientific Revolutions*. 2nd ed. University of Chicago Press, Chicago.

Lindley, D. (1956). On a measure of the information provided by an experiment. In *Annals of Mathematical Statistics*. 27: 987-990.

Peirce, C.S. (1878). The probability of induction. In J. Newman, editor, *The World of Mathematics, II*, 393-396. Simon & Shuster, New York.

Polya, G. (1954). *Patterns of Plausible Reasoning*. Cambridge University Press, Cambridge.

Popper, K.R. (1959). *The Logic of Scientific Discovery*. Hutchinson, London.

Ramsey, F.P. (1931). Truth and probability. In R. Braithwaite, editor, *The Foundations of Mathematics and Other Logical Essays*, 115–162. Kegan Paul, London.

Robertson, B. and Vignaux, G.A. (1993). Probability – the logic of the law. In *Oxford Journal of Legal Studies*. 13: 457-478.

Salmon, W.C. (1990). Rationality and objectivity in science or Tom Kuhn meets Tom Bayes. In C. Wade Savage, editor, *Minnesota Studies in the Philosophy of Science*, 175–204. University of Minnesota Press, Minneapolis.

Shannon, C.E. and Weaver, W. (1949). *The Mathematical Theory of Communication*. University of Illinois Press, Urbana.

Skyrms, B. (1990). *The Dynamics of Rational Deliberation*. Harvard University Press, Cambridge (Mass.).

Taroni, F., Aitken, C.G.G., Garbolino, P. and Biedermann, A. (2006). *Bayesian Networks and Probabilistic Inference in Forensic Science*. Wiley, Chichester.

Teller, P. (1973). Conditionalization and observation. In *Synthese*. 26: 218-258.