City Research Online

City, University of London Institutional Repository

A quantum framework for likelihood ratios

Rachael L. Bond[*] and Thomas C. Ormerod[†]
The School of Psychology, The University of Sussex,
Falmer, East Sussex, BN1 9QH, United Kingdom.

Yang-Hui He[‡]
Department of Mathematics, City University, London, EC1V 0HB, United Kingdom, and
School of Physics, NanKai University, Tianjin, 300071, P.R. China, and and
Merton College, Oxford University, Oxford, OX1 4JD, United Kingdom.
(Dated: August 6, 2015)

The ability to calculate precise likelihood ratios is fundamental to many STEM areas, such as decision-making theory, biomedical science, and engineering. However, there is no assumption-free statistical methodology to achieve this. For instance, in the absence of data relating to covariate overlap, the widely used Bayes' theorem either defaults to the marginal probability driven "naive Bayes' classifier", or requires the use of compensatory expectation-maximization techniques. Equally, the use of alternative statistical approaches, such as multivariate logistic regression, may be confounded by other axiomatic conditions, e.g., low levels of co-linearity. This article takes an information-theoretic approach in developing a new statistical formula for the calculation of likelihood ratios based on the principles of quantum entanglement. In doing so, it is argued that this quantum approach demonstrates: that the likelihood ratio is a real quality of statistical systems; that the naive Bayes' classifier is a special case of a more general quantum mechanical expression; and that only a quantum mechanical approach can overcome the axiomatic limitations of classical statistics.

In recent years, Bayesian statistical research has often been epistemologically driven, guided by de Finetti's famous quote that "probability does not exist"[1]. For example, the "quantum Bayesian" methodology of Caves, Fuchs, & Schack[2] has applied de Finetti's ideas to Bayes' theorem for use in quantum mechanics. In doing so, Caves et al. have argued that statistical systems are best interpreted by methods in which the Bayesian likelihood ratio is seen to be both external to the system and subjectively imposed on it by the observer[3].

From a decision-making perspective the Caves et al. approach is problematic. Bayes' theorem and, in particular, the "naive Bayes' classifier" have been used extensively to interpret information systems and develop normative decision-making models[4]. While subjectivity may play a role in a descriptive model of human decision-making, its use in normative analysis could suggest the presence of a cognitive "homonculus" with the power to influence decision outcomes. Yet at a human scale, for instance, an observer's belief as to the chances of a fair coin landing either "heads" or "tails" has no known effect. Rather, within normative decision-making theory, the "heads:tails" likelihood ratio of 0.5:0.5 is only meaningful when considered as a property of the coin's own internal statistical system rather than as some ephemeral and arbitrary qualia.

However, there are axiomatic difficulties associated with Bayes' theorem, notably its reliance upon the use of marginal probabilities in the absence of structural statistical information, such as estimates of covariate overlap. This has led some researchers to attempt to reconceptualise psychology and decision-making theory using quantum mechanics[5, 6] - an approach with intuitive merit given that both disciplines apply statistical axioms to analyse and interpret probabilistic systems. Yet, despite progress made in this area, the lack of an orthodox Copenhagen-based theoretical counterpoint to Caves et al. has impeded the development of new normative models. It is this knowledge gap which this article aims to fill.

I. THE LIMITS OF BAYES' THEOREM

Bayes' theorem is used to calculate the conditional probability of a statement, or hypothesis, being true given that other information is also true. It is usually written as

$$P(H_i|D) = \frac{P(H_i)P(D|H_i)}{\sum\limits_j P(H_j)P(D|H_j)} \ . \qquad (1)$$

Here, $P(H_i|D)$ is the conditional probability of hypothesis H_i being true given that the information D is true; $P(D|H_i)$ is the conditional probability of D being true if H_i is true; and $\sum\limits_j P(H_j)P(D|H_j)$ is the sum of the probabilities of all hypotheses multiplied by the conditional probability of D being true for each hypothesis[4].

arXiv:1508.00936v1 [math.ST] 4 Aug 2015

	Street A (H_1)	Street B (H_2)
Number of houses (n)	10	10
% blue front door (D)	0.8	0.7
% not blue front door ($\bar{D}$)	0.2	0.3

$$(2)$$

To exemplify using the contingency information in (2), if one wishes to calculate which of two streets is most likely to have a house with a blue front door, then using Bayes' theorem it is trivial to calculate that Street A is most likely with a likelihood ratio of approximately 0.53:0.47,

$$P(H_1|D) = \frac{0.5 \times 0.8}{(0.5 \times 0.8) + (0.5 \times 0.7)} = \frac{8}{15} \approx 0.533 \ ,$$

$$P(H_2|D) = 1 - P(H_1|D) = \frac{7}{15} \approx 0.467 \ ,$$

$$(3)$$

where $P(H_i) = 10/(10 + 10) = 0.5$ for both $i = 1, 2$.

However, difficulties arise in the use of Bayes' theorem for the calculation of likelihood ratios where there are multiple non-exclusive data sets. For instance, if the information in (2) is expanded to include data about the number of houses with a garage (4) then the precise covariate overlap (i.e., $D_1 \cap D_2$) for each street becomes an unknown.

	Street A (H_1)	Street B (H_2)
Number of houses (n)	10	10
% blue front door (D_1)	0.8	0.7
% with garage (D_2)	0.6	0.5

$$(4)$$

All that may be shown is that, for each street, the number of houses with both features forms a range described by (5), where $n(H_i)$ is the total number of houses in street i, $n(D_1|H_i)$ is the total number of houses in street i with a blue front door, and $n(D_2|H_i)$ is the total number of houses in street i with a garage,

$$n(D_1 \cap D_2|H_i) \in$$

$$\begin{cases} \left[n(D_1|H_i) + n(D_2|H_i) - n(H_i) \ , \ldots, \min(n(D_1|H_i), n(D_2|H_i))\right] \\ \qquad \text{if } n(D_1|H_i) + n(D_2|H_i) > n(H_i) \ , \quad \text{or} \\ \left[0 \ , \ldots, \min(n(D_1|H_i), n(D_2|H_i))\right] \\ \qquad \text{if } n(D_1|H_i) + n(D_2|H_i) \le n(H_i) \ . \end{cases}$$

$$(5)$$

Specifically for (4) these ranges equate to

$$n(D_1 \cap D_2|H_1) \in \{4, 5, 6\} \ ,$$
$$n(D_1 \cap D_2|H_2) \in \{2, 3, 4, 5\} \ . \qquad (6)$$

The simplest approach to resolving this problem is to naively ignore any intersection, or co-dependence, of the data and to directly multiply the marginal probabilities[7–9, for example]. Hence, given (4), the likelihood of Street A having the greatest number of houses with both a blue front door and a garage would be calculated as

$$P(H_1|D_1 \cap D_2) = \frac{0.5 \times 0.8 \times 0.6}{(0.5 \times 0.8 \times 0.6) + (0.5 \times 0.7 \times 0.5)}$$
$$\approx 0.578 \ . \qquad (7)$$

Yet, because the data intersect, this probability value is only one of a number which may be reasonably calculated. Alternatives include calculating a likelihood ratio using the mean value μ of the frequency ranges for each hypothesis

$$P(\mu[n(D_1 \cap D_2|H_1)]) = \frac{1}{10} \times \frac{1}{3}(4 + 5 + 6) = 0.5 \ ,$$

$$P(\mu[n(D_1 \cap D_2|H_2)]) = \frac{1}{10} \times \frac{1}{4}(2 + 3 + 4 + 5) = 0.35$$

$$\Rightarrow P(H_1|\mu D_1 \cap D_2) \approx 0.588 \ ; \qquad (8)$$

and taking the mean value of the probability range derived from the frequency range

$$\min P(H_1|D_1 \cap D_2) = \frac{4}{4 + 5} \ ,$$

$$\max P(H_1|D_1 \cap D_2) = \frac{6}{6 + 2}$$

$$\Rightarrow \mu[P(H_1|D_1 \cap D_2)] \approx 0.597 \ . \qquad (9)$$

Given this multiplicity of probability values, it would seem that none of these methods may lay claim to normativity. This problem of covariate overlap has, of course, been previously addressed within statistical literature. For instance, the "maximum likelihood" approach, of Dempster, Laird, & Rubin[10], has demonstrated how an "expectation-maximization" algorithm may be used to derive appropriate covariate overlap measures. Indeed, the mathematical efficacy of this technique has been confirmed by Wu[11]. However, from a psychological standpoint, it is difficult to see how such an iterative methodology could be employed in human decision-making, given mankind's reliance on heuristics rather than cognitively expensive exactitudes[12]. Since there is also little evidence that the naive Bayes' classifier forms any part of the human decision-making process[7], the theoretical advancement of the psychology of decision-making demands a mathematical approach in which covariate overlaps can be automatically, and directly, calculated from contin-

gency data.

II. A QUANTUM MECHANICAL PROOF OF BAYES' THEOREM FOR INDEPENDENT DATA

Previously unconsidered, the quantum mechanical von Neumann axioms would seem to offer the most promise in this regard, since the re-conceptualization of covariate data as a quantum entangled system allows for statistical analysis with few, non-arbitrary assumptions. Unfortunately there are many conceptual difficulties that can arise here. For instance, a Dirac notation representation of (4) as a standard quantum superposition is

$$|\Psi\rangle = \frac{1}{\sqrt{N}}\left[\alpha\left(\sqrt{\frac{1}{3}}\,|4\rangle_{H1} + \sqrt{\frac{1}{3}}\,|5\rangle_{H1} + \sqrt{\frac{1}{3}}\,|6\rangle_{H1}\right)\right.$$
$$\left. + \beta\left(\sqrt{\frac{1}{4}}\,|2\rangle_{H2} + \sqrt{\frac{1}{4}}\,|3\rangle_{H2} + \sqrt{\frac{1}{4}}\,|4\rangle_{H2} + \sqrt{\frac{1}{4}}\,|5\rangle_{H2}\right)\right] .$$
(10)

In this example, (10) cannot be solved since the possible values of $D_1 \cap D_2$ for each hypothesis (6) have been described as equal chance outcomes within a general superposition of H_1 and H_2, with the unknown coefficients α and β assuming the role of the classical Bayesian likelihood ratio.

The development of an alternative quantum mechanical description necessitates a return to the simplest form of Bayes' theorem using the case of exclusive populations H_i and data sets $D, \bar{D}$, such as given in (2). Here, the overall probability of H_1 may be simply calculated as

$$P(H_1) = \frac{n(H_1)}{n(H_1) + n(H_2)} .$$
(11)

The a priori uncertainty in (2) may be expressed by constructing a wave function in which the four data points are encoded as a linear superposition

$$|\Psi\rangle = \alpha_{1,1}\,|H_1 \otimes D\rangle + \alpha_{1,2}\,|H_1 \otimes \bar{D}\rangle$$
$$+ \alpha_{2,1}\,|H_2 \otimes D\rangle + \alpha_{2,2}\,|H_2 \otimes \bar{D}\rangle .$$
(12)

Since there is no overlap between either D and $\bar{D}$ or the populations H_1 and H_2, each datum automatically forms an eigenstate basis with the orthonormal conditions

$$\langle H_1 \otimes D | H_1 \otimes D\rangle = \langle H_1 \otimes \bar{D} | H_1 \otimes \bar{D}\rangle = 1$$
$$\langle H_2 \otimes D | H_2 \otimes D\rangle = \langle H_2 \otimes \bar{D} | H_2 \otimes \bar{D}\rangle = 1$$
$$\text{all other bra-kets } = 0 ,$$
(13)

where the normalization of the wave function demands that

$$\langle\Psi|\Psi\rangle = 1 ,$$
(14)

so that the sum of the modulus squares of the coefficients

$\alpha_{i,j}$ gives a total probability of 1

$$|\alpha_{1,1}|^2 + |\alpha_{1,2}|^2 + |\alpha_{2,1}|^2 + |\alpha_{2,2}|^2 = 1 .$$
(15)

For simplicity let

$$x_1 = P(D|H_1), \ \ y_1 = P(\bar{D}|H_1) ,$$
$$x_2 = P(D|H_2), \ \ y_2 = P(\bar{D}|H_2) ,$$
$$X_1 = P(H_1), \ \ X_2 = P(H_2) .$$
(16)

If the coefficients $\alpha_{i,j}$ from (12) are set as required by (2), it follows that

$$|\alpha_{1,1}|^2 = x_1, \ \ |\alpha_{1,2}|^2 = y_1, \ \ |\alpha_{2,1}|^2 = x_2, \ \ |\alpha_{2,2}|^2 = y_2 ,$$
(17)

so that the normalised wave function $|\Psi\rangle$ is described by

$$|\Psi\rangle = \frac{1}{\sqrt{N}}(\sqrt{x_1}\,|H_1 \otimes D\rangle + \sqrt{y_1}\,|H_1 \otimes \bar{D}\rangle$$
$$+ \sqrt{x_2}\,|H_2 \otimes D\rangle + \sqrt{y_2}\,|H_2 \otimes \bar{D}\rangle)$$
(18)

for some normalization constant N.

The orthonormality condition (14) implies that

$$N = x_1 + y_1 + x_2 + y_2 = X_1 + X_2 ,$$
(19)

thereby giving the full wave function description

$$|\Psi\rangle = \frac{\sqrt{x_1}\,|H_1 \otimes D\rangle + \sqrt{y_1}\,|H_1 \otimes \bar{D}\rangle + \sqrt{x_2}\,|H_2 \otimes D\rangle + \sqrt{y_2}\,|H_2 \otimes \bar{D}\rangle}{\sqrt{X_1 + X_2}} .$$
(20)

If the value of $P(H_1|D)$ is to be calculated, i.e., the property D is observed, then the normalized wave function (12) necessarily collapses to

$$|\Psi'\rangle = \alpha_1\,|H_1 \otimes D_1\rangle + \alpha_2\,|H_2 \otimes D_1\rangle ,$$
(21)

where the coefficients $\alpha_{1,2}$ may be determined by projecting $|\Psi\rangle$ on to the two terms in $|\Psi'\rangle$ using (13), giving

$$\alpha_1 = \langle\Psi'|H_1 \otimes D\rangle = \sqrt{\frac{x_1}{X_1 + X_2}} ,$$

$$\alpha_2 = \langle\Psi'|H_2 \otimes D\rangle = \sqrt{\frac{x_2}{X_1 + X_2}} .$$
(22)

Normalizing (21) with the coefficient N'

$$|\Psi'\rangle = \frac{1}{\sqrt{N'}}\left(\sqrt{\frac{x_1}{X_1 + X_2}}\,|H_1 \otimes D\rangle + \sqrt{\frac{x_2}{X_1 + X_2}}\,|H_2 \otimes D\rangle\right) ,$$
(23)

and using the normalization condition (14), implies that

$$1 = \langle \Psi' | \Psi' \rangle = \frac{1}{N'} \left(\frac{x_1}{X_1 + X_2} + \frac{x_2}{X_1 + X_2} \right)$$

$$\rightarrow N' = \frac{x_1 + x_2}{X_1 + X_2} \ . \tag{24}$$

Thus, after collapse, the properly normalized wave function (23) becomes

$$|\Psi'\rangle = \sqrt{\frac{x_1}{x_1 + x_2}} |H_1 \otimes D\rangle + \sqrt{\frac{x_2}{x_1 + x_2}} |H_2 \otimes D\rangle \ , \tag{25}$$

which means that the probability of observing $|H_1 \otimes D\rangle$ is

$$P(|H_1 \otimes D\rangle) = \left(\sqrt{\frac{x_1}{x_1 + x_2}} \right)^2 = \frac{\alpha_1^2}{\alpha_1^2 + \alpha_2^2} = \frac{x_1}{x_1 + x_2} \ . \tag{26}$$

This is entirely consistent with Bayes' theorem and demonstrates its derivation using quantum mechanical axioms.

III. A QUANTUM MECHANICAL EXPRESSION TO CALCULATE LIKELIHOOD RATIOS, WITH CO-DEPENDENT DATA

Having established the principle of using a quantum mechanical approach for the calculation of simple likelihood ratios with mutually exclusive data (2), it is now possible to consider the general case of n hypotheses and m data (27), where the data are co-dependent, or intersect.

$$\begin{array}{|c|c|c|c|c|}
\hline
 & H_1 & H_2 & \cdots & H_n \\
\hline
D_1 & x_{1,1} & x_{1,2} & \cdots & x_{1,n} \\
\hline
D_2 & x_{2,1} & x_{2,2} & \cdots & x_{2,n} \\
\hline
\vdots & & \vdots & & \vdots \\
\hline
D_m & x_{m,1} & x_{m,2} & \cdots & x_{m,n} \\
\hline
\end{array} \tag{27}$$

Here the contingency table in (27) has been indexed using

$$x_{i,\alpha} \ , \qquad \alpha = 1, 2, \ldots, n; \quad i = 1, 2, \ldots, m \ . \tag{28}$$

While the general wave function remains the same as before, the overlapping data create non-orthonormal inner

products which can be naturally defined as

$$\langle H_\alpha \otimes D_i | H_\beta \otimes D_j \rangle = c_{ij}^\alpha \delta_{\alpha\beta} \ , \ \ c_{ij}^\alpha = c_{ji}^\alpha \in \mathbb{R} \ , c_{ii}^\alpha = 1 \ , \tag{29}$$

Assuming, for simplicity, that the overlaps c_{ij}^α are real, then there is a symmetry in that $c_{ij}^\alpha = c_{ji}^\alpha$ for each α. Further, for each α and i, the state is normalized, i.e., $c_{ii}^\alpha = 1$. The given independence of the hypotheses H_α also enforces the Kroenecker delta function, $\delta_{\alpha\beta}$.

The Hilbert space V spanned by the kets $|H_\alpha \otimes D_i\rangle$ is mn-dimensional and, because of the independence of H_α, naturally decomposes into the direct sum (30) with respect to the inner product, thereby demonstrating that the non-orthonormal conditions are the direct sum of m vector spaces V^α:

$$V = \text{Span}(\{|H_\alpha \otimes D_i\rangle\}) = \bigoplus_{\alpha=1}^n V^\alpha \ , \qquad \dim V^\alpha = m \ . \tag{30}$$

Since the inner products are non-orthonormal, each V^α must be individually orthonormalised. Given that V splits into a direct sum, this may be achieved for each subspace V^α by applying the Gram-Schmidt algorithm[13] to $\{|H_\alpha \otimes D_i\rangle\}$ of V. Consequently, the orthonormal basis may be defined as

$$|K_i^\alpha\rangle = \sum_{k=1}^n A_{i,k}^\alpha |H_\alpha \otimes D_k\rangle \ , \qquad \langle K_i^\alpha | K_j^\alpha \rangle = \delta_{ij} \ , \tag{31}$$

for each $\alpha = 1, 2, \ldots, n$ with $m \times m$ matrices $A_{i,k}^\alpha$, for each α.

Substituting the inner products (29) gives

$$\sum_{k,k'=1}^m A_{ik}^\alpha A_{jk'}^\alpha c_{kk'}^\alpha = \delta_{ij} \quad \forall \alpha = 1, 2, \ldots, n \ . \tag{32}$$

The wave-function may now be written as a linear combination of the orthonormalised kets $|K_i^\alpha\rangle$ with the coefficients b_i^α, and may be expanded into the $|H_\alpha \otimes D_i\rangle$ basis using (31), i.e.,

$$|\Psi\rangle = \sum_{\alpha,i} b_i^\alpha |K_i^\alpha\rangle = \sum_{\alpha,i,k} b_i^\alpha A_{ik}^\alpha |H_\alpha \otimes D_k\rangle \ . \tag{33}$$

As with (17) from earlier, the coefficients in (33) should be set as required by the contingency table

$$\sum_i b_i^\alpha A_{i,k}^\alpha = \sqrt{x_{k\alpha}} \ , \tag{34}$$

where, to solve for the b-coefficients, (32) may be used to invert

$$\sum_{k,k'} \sum_i b_i^\alpha A_{ik}^\alpha A_{jk'}^\alpha c_{kk'}^\alpha = \sum_{k,k'} \sqrt{x_{k\alpha}} A_{jk'}^\alpha c_{k'k}^\alpha \ , \tag{35}$$

giving

$$b_j^\alpha = \sum_{k,k'} \sqrt{x_{k\alpha}} A_{jk'}^\alpha c_{kk'}^\alpha \ . \qquad (36)$$

Having relabelled the indices as necessary, a back-substitution of (34) into the expansion (33) gives

$$|\Psi\rangle = \sum_{\alpha,i,k} b_i^\alpha A_{i,k}^\alpha |H_\alpha \otimes D_k\rangle = \sum_{\alpha,k} \sqrt{x_{k\alpha}} |H_\alpha \otimes D_k\rangle \ , \qquad (37)$$

which is the same as having simply assigned each ket's coefficient to the square root of its associated entry in the contingency table.

The normalization factor for $|\Psi\rangle$ is simply $1/\sqrt{N}$, where N is the sum of the squares of the coefficients b of the orthonormalised bases $|K_i^\alpha\rangle$,

$$N = \sum_{i,\alpha} (b_i^\alpha)^2 = \sum_{i,\alpha} b_i^\alpha \left(\sum_{k,k'} \sqrt{x_{k\alpha}} A_{k',i}^\alpha c_{kk'}^\alpha \right)$$
$$= \sum_{k,k',\alpha} \sqrt{x_{k\alpha} x_{k'\alpha}} c_{kk'}^\alpha \ . \qquad (38)$$

Thus, the final normalized wave function is

$$|\Psi\rangle = \frac{\sum\limits_{\alpha,k} \sqrt{x_{k\alpha}} |H_\alpha \otimes D_k\rangle}{\sqrt{\sum\limits_{i,j,\alpha} \sqrt{x_{i\alpha} x_{j\alpha}} c_{ij}^\alpha}} \ , \qquad (39)$$

where α is summed from 1 to n, and i,j are summed from 1 to m. Note that, in the denominator, the diagonal term $\sqrt{x_{i\alpha} x_{j\alpha}} c_{ij}^\alpha$, which occurs whenever $i = j$, simplifies to $x_{i\alpha}$ since $c_{ii}^\alpha = 1$ for all α.

From (39) it follows that, exactly in parallel to the non-intersecting case, if all properties D_i are observed simultaneously, the probability of any hypothesis H_α, for a fixed α, is

$$P(H_\alpha | D_1 \cap D_2 \ldots \cap D_m) = \frac{\sum\limits_i (b_i^\alpha)^2}{\sum\limits_{i,\alpha} (b_i^\alpha)^2} = \frac{\sum\limits_{i,j} \sqrt{x_{i\alpha} x_{j\alpha}} c_{ij}^\alpha}{\sum\limits_{i,j,\alpha} \sqrt{x_{i\alpha} x_{j\alpha}} c_{ij}^\alpha} \ , \qquad (40)$$

In the case of non-even populations for each hypothesis (i.e., non-even priors), each element within (40) should be appropriately weighted.

IV. EXAMPLE SOLUTION

Returning to the problem presented in the contingency table (4), it is now possible to calculate the precise probability for a randomly selected house with the properties of "blue front door" and "garage" belonging to Street A (H_1). For this 2×2 matrix, recalling from (29) that

$c_{ii}^\alpha = 1$ and $c_{ij}^\alpha = c_{ji}^\alpha$, the general expression (40) may be written as

$$P(H_1|D_1 \cap D_2) = \frac{\sum\limits_{i,j=1}^{2} \sqrt{x_{i,1} x_{j,1}} c_{ij}^1}{\sum\limits_{i,j=1}^{2} \sum\limits_{\alpha=1}^{2} \sqrt{x_{i\alpha} x_{j\alpha}} c_{ij}^\alpha}$$

$$= \frac{\sqrt{x_{1,1}^2} c_{1,1}^1 + \sqrt{x_{2,1}^2} c_{2,2}^1 + \sqrt{x_{1,1} x_{2,1}} c_{1,2}^1 + \sqrt{x_{2,1} x_{1,1}} c_{2,1}^1}{\sum\limits_{\alpha=1}^{2} \sqrt{x_{1,\alpha}^2} c_{1,1}^1 + \sqrt{x_{2,\alpha}^2} c_{2,2}^1 + \sqrt{x_{1,\alpha} x_{2,\alpha}} c_{1,2}^1 + \sqrt{x_{2,\alpha} x_{1,\alpha}} c_{2,1}^1}$$

$$= \frac{x_1 + y_1 + 2c_1\sqrt{x_1 y_1}}{x_1 + x_2 + y_1 + y_2 + 2c_1\sqrt{x_1 y_1} + 2c_2\sqrt{x_2 y_2}} \ , \qquad (41)$$

where, adhering to the earlier notation (16),

$$x_1 = x_{1,1} = P(D_1|H_1), \ y_1 = x_{2,1} = P(D_2|H_1) \ ,$$
$$x_2 = x_{1,2} = P(D_1|H_2), \ y_2 = x_{2,2} = P(D_2|H_2) \ ,$$
$$X_1 = P(H_1), \ X_2 = P(H_2) \ , \qquad (42)$$

and, for brevity, $c_1 := c_{1,2}^1$, $c_2 := c_{1,2}^2$. For simplicity, $P(H_i|D_1 \cap D_2)$ will henceforth be denoted as P_i. Implementing (41) is dependent upon deriving solutions for the yet unknown expressions c_i, $i = 1, 2$ which govern the extent of the intersection in (29). This can only be achieved by imposing reasonable constraints upon c_i which have been inferred from expected behaviour and known outcomes, i.e., through the use of boundary values and symmetries. Specifically, these constraints are:

Data dependence.: The expressions c_i must, in some way, be dependent upon the data given in the contingency table, i.e.,

$$c_1 = c_1(x_1, y_1, x_2, y_2; X_1, X_2) \ ,$$
$$c_2 = c_2(x_1, y_1, x_2, y_2; X_1, X_2) \ . \qquad (43)$$

Probability.: The calculated values for P_i must fall between 0 and 1. Since x_i and y_i are positive, it suffices to take

$$-1 < c_i(x_1, y_1, x_2, y_2) < 1 \ . \qquad (44)$$

Complementarity.: The law of total probability dictates that

$$P_1 + P_2 = 1 \ , \qquad (45)$$

which can easily be seen to hold.

Symmetry.: The exchanging of rows within the contingency tables should not affect the calculation of P_i. In other words, for each $i = 1, 2$, P_i is invariant under $x_i \leftrightarrow y_i$. This constraint implies that

$$c_i(x_1, y_1, x_2, y_2) = c_i(y_1, x_1, y_2, x_2) \ . \qquad (46)$$

Equally, if the columns are exchanged then P_i must

map to each other, i.e., for each $i = 1, 2$ then $P_1 \leftrightarrow P_2$ under $x_1 \leftrightarrow x_2$, $y_1 \leftrightarrow y_2$ which gives the further constraint that

$$c_1(x_1, y_1, x_2, y_2) = c_2(x_2, y_2, x_1, y_1) \ . \qquad (47)$$

Known values.: There are a number of contingency table structures which give rise to a known probability, i.e.,

	H_1	H_2
D_1	1	1
D_2	m	n

$\rightarrow \qquad P_1 = \dfrac{m}{m+n}$

	H_1	H_2
D_1	m	n
D_2	1	1

$\rightarrow \qquad P_1 = \dfrac{m}{m+n}$

	H_1	H_2
D_1	n	m
D_2	m	n

$\rightarrow \qquad P_1 = \dfrac{1}{2}$

	H_1	H_2
D_1	n	n
D_2	m	m

$\rightarrow \qquad P_1 = \dfrac{1}{2}$

	H_1	H_2
D_1	m	m
D_2	m	m

$\rightarrow \qquad P_1 = \dfrac{1}{2} \ , \qquad (48)$

where m, n are positively valued probabilities. For such contingency tables the correct probabilities should always be returned by c_i. Applying this principle to (41) gives the constraints

$$\frac{m}{m+n} = \frac{2c_1(m,1,n,1)\sqrt{m} + m + 1}{2c_1(m,1,n,1)\sqrt{m} + 2c_2(m,1,n,1)\sqrt{n} + m + n + 2} \ , \qquad (49)$$

$$\frac{1}{2} = \frac{2c_1(n,m,m,n)\sqrt{m}\sqrt{n} + m + n}{2c_1(n,m,m,n)\sqrt{m}\sqrt{n} + 2c_2(n,m,m,n)\sqrt{m}\sqrt{n} + 2m + 2n} \ , \qquad (50)$$

$$\frac{1}{2} = \frac{2c_1(n,m,n,m)\sqrt{m}\sqrt{n} + m + n}{2c_1(n,m,n,m)\sqrt{m}\sqrt{n} + 2c_2(n,m,n,m)\sqrt{m}\sqrt{n} + 2m + 2n} \ . \qquad (51)$$

Non-homogeneity.: Bayes' theorem returns the same probability for any linearly scaled contingency ta-

bles, e.g.,

$$x_1 \to 1.0, \ y_1 \to 1.0, \ x_2 \to 1.0, \ y_2 \to 0.50 \Rightarrow P_1 \approx 0.667 \ , \qquad (52)$$
$$x_1 \to 0.5, \ y_1 \to 0.5, \ x_2 \to 0.5, \ y_2 \to 0.25 \Rightarrow P_1 \approx 0.667 \ . \qquad (53)$$

While homogeneity may be justified for conditionally independent data, this is not the case for intersecting, co-dependent data since the act of scaling changes the nature of the intersections and the relationship between them. This may be easily demonstrated by taking the possible value ranges for (52) and (53), calculated using (5), which are

$$\text{Eq. (52)} \Rightarrow (D_1 \cap D_2)|H_1 = \{1\} \ ,$$
$$(D_1 \cap D_2)|H_2 = \{0.5\} \ ,$$

$$\text{Eq. (53)} \Rightarrow (D_1 \cap D_2)|H_1 = \{0.0 \ldots 0.5\} \ ,$$
$$(D_1 \cap D_2)|H_2 = \{0.0 \ldots 0.25\} \ . \qquad (54)$$

The effect of scaling has not only introduced uncertainty where previously there had been none, but has also introduced the possibility of 0 as a valid answer for both hypotheses. Further, the spatial distance between the hypotheses has also decreased. For these reasons it would seem unreasonable to assert that (52) and (53) share the same likelihood ratio.

Using these principles and constraints it becomes possible to solve c_i. From the principle of symmetry it follows that

$$c_1(n,m,m,n) = c_2(m,n,n,m) = c_2(n,m,m,n) \ ,$$
$$c_1(n,m,n,m) = c_2(m,n,n,m) = c_2(n,m,n,m) \ , \qquad (55)$$

and that the equalities (50), (51) for $P_i = 0.5$ automatically hold. Further, (49) solves to give

$$c_2(m,1,n,1) = \frac{2\sqrt{mn}c_1(m,1,n,1) - m + n}{2m\sqrt{n}} \ , \qquad (56)$$

which, because $c_1(n,1,m,1) = c_2(m,1,n,1)$, finally gives

$$c_1(n,1,m,1) = \frac{2\sqrt{mn}c_1(m,1,n,1) - m + n}{2m\sqrt{n}} \ . \qquad (57)$$

Substituting $g(m,n) := \sqrt{n}c_1(m,1,n,1)$ transforms (57) into an anti-symmetric bivariate functional equation in m, n,

$$g(m,n) - g(n,m) = \frac{m}{2\sqrt{mn}} - \frac{n}{2\sqrt{mn}} \ , \qquad (58)$$

whose solution is $g(m,n) = \frac{m}{2\sqrt{mn}}$.

This gives a final solution for the coefficients $c_{1,2}$ of

$$c_1(x_1, y_1, x_2, y_2) = \frac{\sqrt{x_1 y_1}}{2 x_2 y_2} \ ,$$

$$c_2(x_1, y_1, x_2, y_2) = \frac{\sqrt{x_2 y_2}}{2 x_1 y_1} \ . \qquad (59)$$

Thus, substituting (59) into (41) gives the likelihood ratio expression of,

$$P(H_1 | D_1 \cap D_2) = \frac{\frac{x_1 y_1}{x_2 y_2} + x_1 + y_1}{\frac{x_1 y_1}{x_2 y_2} + x_1 + y_1 + \frac{x_2 y_2}{x_1 y_1} + x_2 + y_2} \ . \qquad (60)$$

Given that the population sizes of H_1 and H_2 are the same, no weighting of the elements needs to take place. Hence, the value of $P(H_1 | D_1 \cap D_2)$ for (4) may now be calculated to be

$$P(H_1 | D_1 \cap D_2) \approx 0.5896 \ . \qquad (61)$$

V. DISCUSSION

One of the greatest obstacles in developing any statistical approach is demonstrating correctness. This formula is no different in that respect. If correctness could be demonstrated then, a priori, there would be an appropriate existing method which would negate the need for a new one. All that may be hoped for in any approach is that it generates appropriate answers when they are known, reasonable answers for all other cases, and that these answers follow logically from the underlying mathematics.

However, what is clear is that the limitations of the naive Bayes' classifier render any calculations derived from it open to an unknown margin of error. Given the importance of accurately deriving likelihood ratios this is troubling. This is especially true when these calculations are used to describe normative psychological theories [7] from which inferences are drawn as to the failure of human logic.

As a quantum mechanical methodology this result is able to calculate accurate, iteration free, likelihood ratios which fall beyond the scope of existing statistical techniques, and offers a new theoretical approach within cognitive psychology. Further, through the addition of a Hamiltonian operator to introduce time-evolution, it can offer likelihood ratios for future system states with appropriate updating of the contingency table. In contrast, Bayes' theorem is unable to distinguish directly between time-dependent and time-independent systems. This

may lead to situations where the process of contingency table updating results in the same decisions being made repeatedly with the appearance of an ever increasing degree of certainty. Indeed, from (26), it would seem that the naive Bayes' classifier is only a special case of a more complex quantum mechanical framework, and may only be used where the exclusivity of data is guaranteed.

The introduction of a Hamiltonian operator, and a full quantum dynamical formalism, is in progress, and should have profound implications for cognitive psychology. Inevitably, such a formalism will require a sensible continuous classical limit. In other words, the final expressions for the likelihood ratios should contain a parameter, in some form of $\hbar$, which, when going to 0, reproduces a classically known result. For example, the solutions to (59) could be moderated as

$$c_1(x_1, y_1, x_2, y_2) = \frac{\sqrt{x_1 y_1}}{2 x_2 y_2}(1 - \exp(-\hbar)) \ ,$$

$$c_2(x_1, y_1, x_2, y_2) = \frac{\sqrt{x_2 y_2}}{2 x_1 y_1}(1 - \exp(-\hbar)) \ , \qquad (62)$$

so that in the limit of $\hbar \to 0$, the intersection parameters, c_1 and c_2, vanish to return the formalism to the classical situation of independent data.

VI. CONCLUSION

This article has demonstrated both theoretically, and practically, that a quantum mechanical methodology can overcome the axiomatic limitations of classical statistics in respect to their application within cognitive psychology. In doing so, it challenges the orthodoxy of de Finetti's epistemological approach to statistics by demonstrating that it is possible to derive "real" likelihood ratios from information systems without recourse to arbitrary and subjective evaluations.

While further theoretical development work needs to be undertaken, particularly with regards to the application of these mathematics in other domains, it is hoped that this article will help advance the debate over the nature and meaning of statistics, particularly with respect to cognitive psychology and decision-making theory.

A. Citations and References

[1] B. de Finetti, *Theory of probability: A critical introductory treatment* (Wiley, New York, New York, 1974).

[2] C. M. Caves, C. A. Fuchs, and

R. Schack, Physical Review A **66**, 062111 (2002), arXiv:0206110 [quant-ph].

[3] C. G. Timpson, Studies in History and Philosophy of Science Part B: Studies in History and Philosophy of Modern Physics **39**, 579 (2

[4] M. Oaksford and N. Chater, *Bayesian rationality. The probabilistic approach to human reasoning* (Oxford University Press, Oxford, England, 2007).

[5] A. Khrennikov, BioSystems **95**, 179 (2009).

[6] J. R. Busemeyer and P. D. Bruza, *Quantum models of cognition and decision* (Cambridge University Press, Cambridge, England, 2012).

[7] M. E. Doherty, C. R. Mynatt, R. D. Tweney, and M. D. Schiavo, Acta Psychologica **43**, 111 (1979).

[8] Feeney, A. and Evans, J. and Venn, S., Thinking & Reasoning **14**, 209 (2008).

[9] and, Thinking & Reasoning **39**, 579 (2

[10] A. P. Dempster, N. M. Laird, and D. B. Rubin, Journal of the Royal Statistical Society. Series B (Statistical Method

[11] C. F. J. Wu, The Annals of Statistics **11**, 95 (1983).

[12] D. Kahneman, *Thinking, fast and slow* (Farrar, Straus & Giroux, New York, New York, 2011).

[13] G. Strang, *Linear algebra and its applications*, 2nd ed. (Academic Press, New York, New York, 1980).